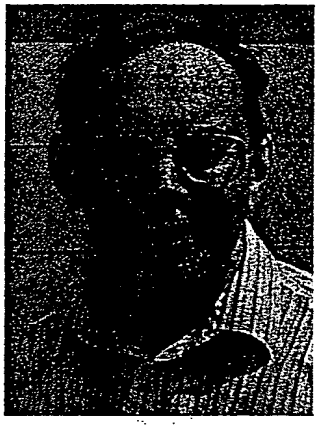
Genetic Contributions to Understanding Polyketide Synthases

David A. Hopwood

John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK

Received January 24, 1997 (Revised Manuscript Received April 23, 1997)

Contents



David Hopwood was born in 1933 in Kinver, England. He received his bachelor and Ph.D. degrees from the University of Cambridge, where he was an assistant lecturer in Botany for five years before moving to the Department of Genetics in the University of Glasgow in 1961. Since 1968 he has been John Innes Professor of Genetics in the University of East Anglia, Norwich, and head of the Genetics Department at the John Innes Centre. From the beginning of his research career he has pioneered studies of the genetics of *Streptomyces*, a member of the group of filamentous, Gram-positive soil bacteria called the actinomycetes that are preeminent producers of polyketide and other antibiotics. Over the last 10 years or so he has harnessed this genetic system to help to illuminate the mechanisms of polyketide biosynthesis.

I. Introduction

Over the last eight years or so, genetic techniques have spearheaded significant advances in our understanding of the structure and mechanisms of polyketide synthases (PKSs). Much insight into these fascinating multifunctional enzymes had already been obtained by chemical and biochemical approaches, including the establishment of a mechanistic relationship between polyketide and fatty acid biosynthesis, in which the carbon backbones of the molecules are assembled by the successive condensation of small acyl units.[1] Severe difficulties had, however, been met in trying to understand a key issue in the field, namely how the enzymes are "programmed". This term has been introduced to describe control of the variables that determine the structure of the product of a specific PKS (Figure 1). These variables are choice of starter unit; choice of the nature and number of the chain extender units; control of the reductive cycle on the β-keto group of the growing carbon chain, which in turn determines the keto, hydroxyl, enoyl, or methylene functionality at each alternate carbon atom; stereochemistry of hydroxyl and alkyl side groups; and pattern of cyclization of the nascent carbon chain. The programming problem is the aspect of polyketide biosynthesis on which genetic studies have had the largest impact. First, the mere cloning and sequencing of the structural genes for a variety of PKSs established the number and primary structures of the

S0009-2665(96)00034-9 CCC: $28.00

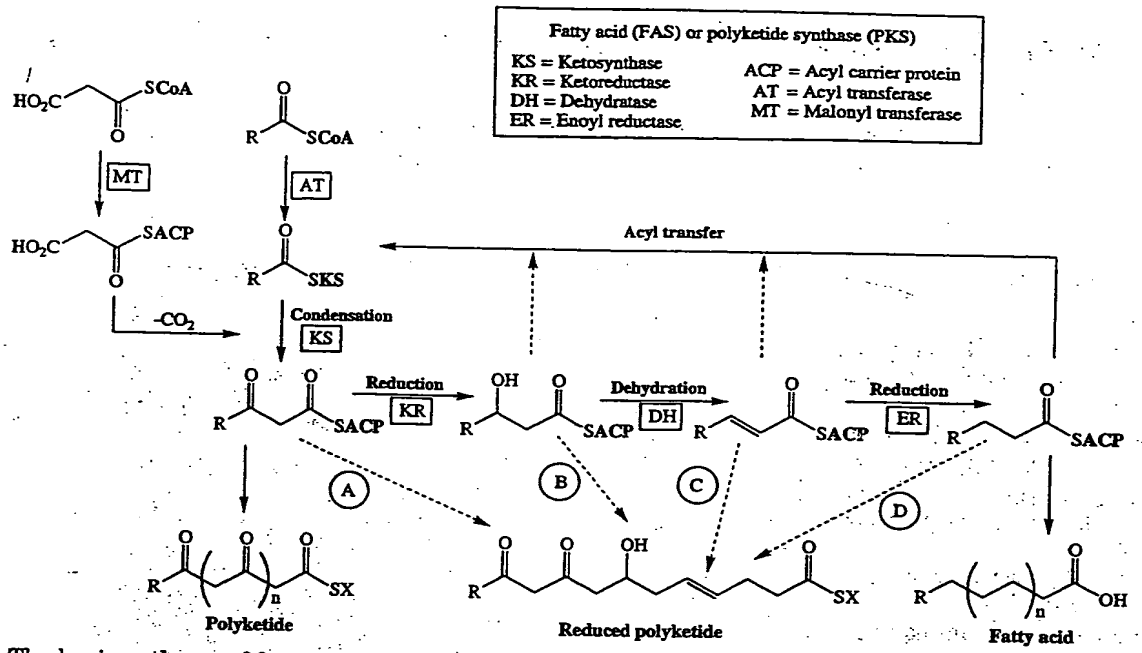


Figure 1. The basic pathway of fatty acid and polyketide biosynthesis, showing the roles of the various activities carried out by the subunits or domains of the fatty acid or polyketide synthase (box). A–D represent the alternative versions of the reductive cycle that lead to keto, hydroxy, enoyl, or methylene functionality, respectively, at specific β-carbons during the assembly of reduced polyketides. Note that the starter unit may be acetyl coenzyme A (CoA) (where R = $-CH_3$) or an alternative CoA ester for all classes of synthase, while the chain extender unit is malonyl CoA (as shown) for the synthesis of fatty acids and aromatic polyketides, but varies for reduced polyketides: incorporation of propionate or butyrate residues (from methylmalonyl CoA or ethylmalonyl CoA chain extenders) produces methyl or ethyl side chains in the polyketide product. (From ref 2, with permission of *Chemistry and Industry*. Copyright 1995 Society of Chemical Industry. Figure kindly provided by T. J. Simpson.)

protein subunits that make up any particular synthase (Figure 2). Examples were found of PKSs resembling each of the classical classes of fatty acid synthases (FASs): type I FAS, characteristic of fungi and vertebrates, in which the catalytic sites for the various steps in the biosynthesis—acyl transferase (AT), ketosynthase (KS), acyl carrier protein (ACP), ketoreductase (KR), dehydratase (DH), enoylreductase (ER) (Figure 1)—are carried as domains along the length of multifunctional proteins; and type II FAS, characteristic of bacteria and plants, in which each catalytic site is carried on a separate protein subunit.[3] Not surprisingly (at least in retrospect) the first example of a fungal PKS had a type I organization, while the first bacterial PKSs to be studied—for members of the aromatic family of polyketides from the actinomycetes—turned out to have a type II structure. In contrast, and totally unexpectedly, the gene sequences for PKSs for the macrolide polyketides of actinomycetes revealed not only a type I organization, hitherto known only in eukaryotes, but the presence of multiple sets, or modules, of active sites. Each module resembled a vertebrate type I FAS, and the whole PKS consisted of a number of modules equal to the number of rounds of condensation required to build the polyketide product. Just as the structure of the double helix had immediately suggested a basis for the faithful replication of DNA,[14] so too the primary structure of the macrolide PKS provided a compelling hypothesis for the programming of the enzyme. On this hypothesis, the program was hard-wired in the gene sequence and was expressed in the encoded proteins as a series of active sites appropriately arranged in relation to each other. The polyketide would be built on an assembly line represented by these active sites, with a loading module at the start; followed by the appropriate number of chain-extending modules in correct sequence and each carrying the relevant complement of reductive sites; and an end domain represented by a thioesterase. The thioesterase would hydrolyze the bond between the completed polyketide and the 4′-phosphopantetheine prosthetic group on the acyl carrier domain of the last chain-extending module, just as in a vertebrate FAS. This hypothesis had to be experimentally verified, but it arose from the sequence itself. In contrast, insight into the quite different programming mechanism of the nonmodular PKSs, in which there is only one catalytic site of each type, and this has to act iteratively to build and modify the polyketide chain, came, not directly from the sequences of the genes, but from the results of their experimental manipulation.

In this Review I have taken a somewhat historical view of the understanding of the PKSs acquired via genetic approaches, rather than attempting an exhaustive review of developments in the field over a specific recent time period. Most of the genetic work with both the type II and the modular PKSs has concerned the streptomycetes and their near relatives among the actinomycetes, and this occupies the bulk of the article in sections II and III, respectively. A lesser amount of genetic work has involved the type I PKSs of filamentous fungi, reviewed in section IV. Outside of microbial systems, work on higher plants has provided a wealth of knowledge about a group of PKSs—the chalcone and stilbene synthases—that probably represent a separate line of evolution from

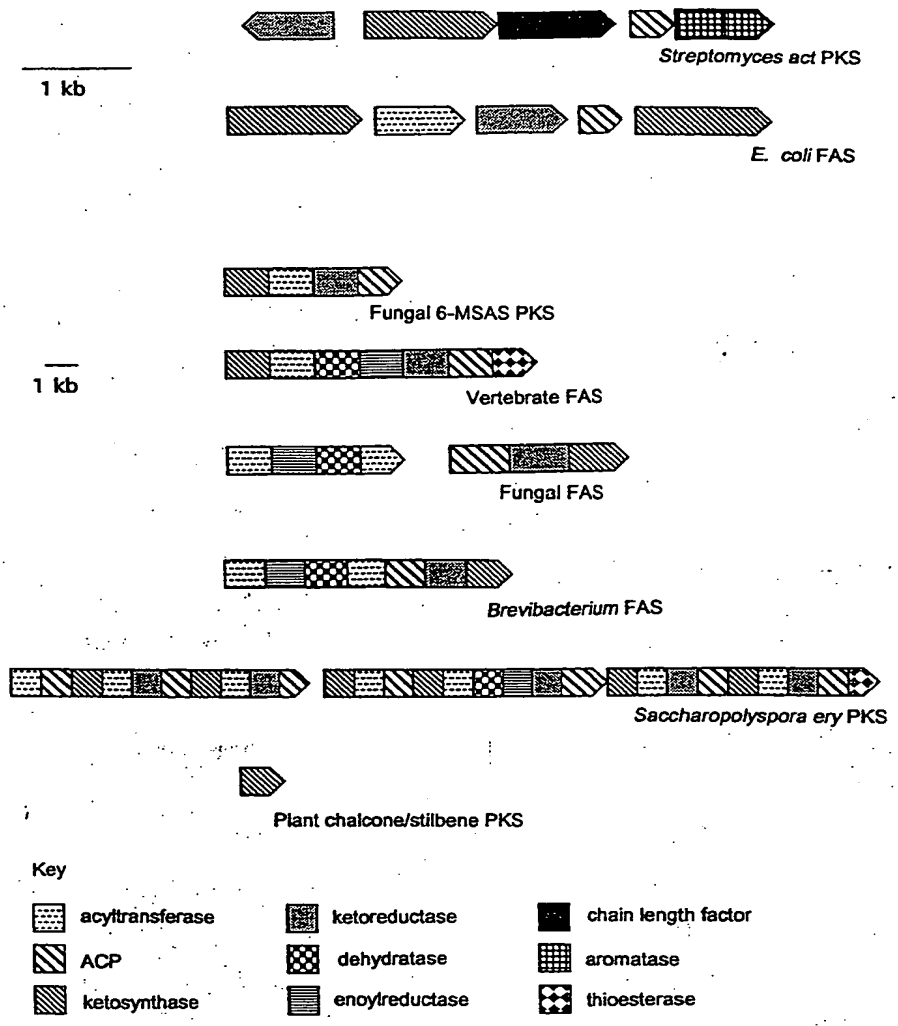


Figure 2. Architecture of fatty acid synthases (FAS) and polyketide synthases (PKS) deduced from the gene sequences. (Note the different scale for the type II synthases (top) and the type I and chalcone/stilbene synthases (bottom).) References to PKS gene sequences for *Streptomyces*, *Saccharopolyspora*, fungi, and plants are in sections II, III, IV, and V, respectively, where these genes are discussed in detail. For the FASs the references are: *E. coli*;[4] vertebrates;[5-7] fungi;[8-12] *Brevibacterium*.[13] The organization and possible evolutionary relationships between the various types of synthase are discussed later, in section VI.

the other PKSs and all of the known FASs, and this is discussed in section V. Section VI is devoted to some evolutionary speculations. The genes that encode the "tailoring" steps that typically follow the building and folding of the polyketide carbon chain are not discussed; genetics has greatly illuminated the nature and mechanisms of many of the enzymes that catalyze these post-PKS steps also, but these studies merge into the whole field of metabolic pathway analysis and manipulation that is not uniquely associated with the polyketides. Other recent reviews of polyketide biosynthesis, some of which take a more chemical stance than my article, include ref 2 and 15–22.

II. Aromatic Polyketide Synthases from Streptomyces Species and Related Actinomycetes

A. Cloning of the Genes

Soon after methods for gene cloning in *Streptomyces* species were published in 1980[23-25] it became possible to isolate genes for antibiotic biosynthesis by a variety of procedures.[26,27] One of the first approaches involved the shotgun cloning of random fragments of DNA from a wild-type strain into a mutant blocked at a step in the biosynthesis and looking for a restoration of antibiotic production. A second early approach depended on emerging evidence for close linkage between genes for self-resistance to an antibiotic and one or more of the biosynthetic genes. From these early results a most important generalization soon emerged: that in streptomycetes, and by implication in other bacteria too, all of the biosynthetic genes needed to make a particular antibiotic from primary metabolites occur together in a single cluster, and that one or more genes for antibiotic self-resistance are also to be found there.[28] A striking early demonstration of this was the cloning of the entire cluster of genes (the *act* genes) for biosynthesis of the pigmented benzoisochromanequinone polyketide actinorhodin (1, Figure 3) on a 35 kb fragment of chromosomal DNA from the producer, *Streptomyces coelicolor* A3(2), and their

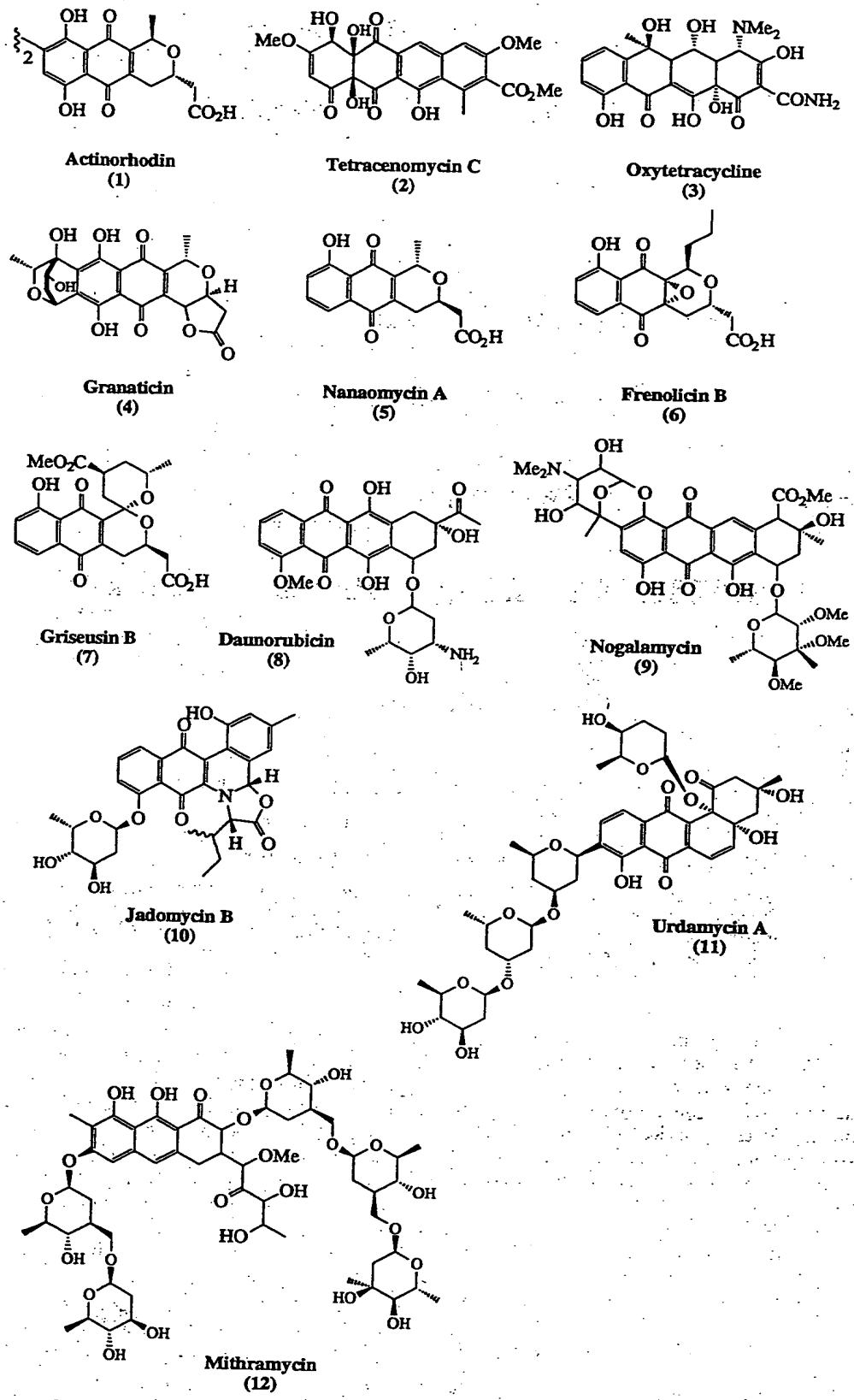


Figure 3. Structures of some aromatic polyketide antibiotics produced by type II PKSs in *Streptomyces* species. (Figure kindly drawn by C. Surti.)

expression in the actinorhodin-sensitive and nonproducing *Streptomyces parvulus*, which thereupon produced actinorhodin without killing itself.[29] Establishment of this generalization—which has continued to hold in all examples known to date—provided two general routes to the isolation of potentially complete sets of genes for the biosynthesis of polyketides by actinomycetes: (1) cloning of a library of fairly large

Table 1. Cloning and Sequencing of Aromatic PKS Gene Clusters From Actinomycetes[a]

host	polyketide	PKS genes	cloning strategy	nucleotide sequence accession number	evidence for cloning correct genes[b]	ref(s)[c]
S. coelicolor	actinorhodin (1)	*act*	Complementation	M19536 X63449	1,2,3	29,37,38
S. rimosus	oxytetracycline (3)	*otc*	Resistance/ complementation	Z25538	1,3	31,39–41
S. glaucescens	tetracenomycin (2)	*tcm*	Complementation	M80674	1,2,3	30,42–44
S. violaceoruber	granaticin (4)	*gra*	*act* probe	X16144	1,2	33,45
S. coelicolor	spore pigment	*whiE*	Complementation/ *act* probe	X55942	1,2	33,35
S. peucetius	daunorubicin	*dps*	*act/tcm* probes	L35560	1,2,3	46,47
S. cinnamonensis	unknown	*mon*	*act* probe	Z11511		34
S. halstedii	spore pigment	*sch*	*act* probe	L05390	1,2	48
S. curacoi	probably spore pigment	*cur*	*act* probe	M33704		36
Sac. hirsuta	unknown	*hir*	*act* probe	M98258		49
S. roseofulvus	nanaomycin (5)/frenolicin (6)	*fren*	*act* probe	L26338	4	50
S. griseus	griseusin (7)	*gris*	*act* probe	X77865	2	51
S. venezuelae	jadomycin (10)	*jad*	*act* probe	L33245	2	52
S. sp. C5	daunorubicin (8)	*dau*	*act* probe	L34880	1	53
Kib. aridum	unknown	*ard*	*act* probes	L24518		54
S. fradiae	urdamycin (11)	*urd*	*tcm/act* probes	X87093		55
S. nogalater	nogalamycin (9)	*sno*	*act* probe	Z48262	1,2,3	56,57
S. argillaceus	mithramycin (12)	*mtm*	*act* probe	X89899	2	58

[a] Other PKS gene sets that have been cloned but not (fully) sequenced include those encoding the PKSs for aclacinomycin,[59,60] kalafungin,[61] elloramycin,[62] PD117420, and tetrangulol.[63] [b] Key: 1, complementation of pathway-blocked mutants; 2, gene disruption; 3, production of antibiotic after transfer of cloned genes to *S. lividans*; 4, production of relevant compounds in recombinant strains. [c] These include references to the cloning and/or sequencing of the PKS genes.

of the pigment was (and still is) unknown, but the sequence of the complementing DNA left little doubt that it represented PKS genes for an aromatic polyketide (see below), and indeed the *act* and *whiE* PKS genes cross-hybridized, so that the *act*I probe revealed two bands when hybridized to *S. coelicolor* genomic DNA.[33] Probably reflecting a similar situation, use of the *act* probes to try to isolate the PKS for the simple aromatic polyketide curamycin from *Streptomyces curacoi* yielded DNA that may well encode the PKS for a polyketide spore pigment rather than the antibiotic, although this remains unproven.[36]

In spite of these setbacks, the *act* probes have been instrumental in the isolation of sets of genes that encode the PKSs for a variety of aromatic metabolites (Figure 3), so that by now at least 18 have been cloned and sequenced (Table 1), including further examples of benzoisochromanequinones, nanaomycin (5), frenolicin (6), and griseusin (7); and anthracyclines, daunorubicin (8), the closely related doxorubicin (adriamycin) and nogalamycin (9); two angucyclines, jadomycin (10) and urdamycin (11); and the aureolic acid derivative mithramycin (12). Some others have been cloned but not yet sequenced (Table 1, footnote). Table 1 also lists the classes of evidence adduced to prove involvement of each gene set in biosynthesis of the target polyketide. The most general of these are (1) complementation of *bona fide* blocked mutants by the cloned PKS DNA to restore production of the relevant polyketide; (2) disruption, using fragments of the cloned DNA, of the corresponding genes in the wild-type strain, to generate a nonproducing phenotype; and (3) transfer of a set of antibiotic biosynthetic genes, including the presumptive PKS genes, to a heterologous host (usually *S. lividans*), resulting in production of the relevant antibiotic (or a precursor of it).

It is now clear that, although all PKSs (except the chalcone/stilbene family) may well share the same evolutionary origin, the synthases for the aromatic class and the "complex" or "reduced" class of polyketides (most famously represented by the various kinds of macrolides) appear to be too far diverged from each other for significant cross-hybridization at the DNA level to occur. Thus the *act* probes are in general not useful for isolating genes for the modular family of PKSs (but see under soraphen, section III.B.7). For these, a different series of probes, usually segments of DNA encoding various domains of the erythromycin PKS, are effective (see section III). Both classes of PKS genes appear to have diverged too far from any FAS genes for cross-hybridization at the DNA level to occur, so hybridization of PKS probes with FAS genes is not usually a problem.

B. Architecture of the Genes

The first clue to the architecture of a bacterial polyketide synthase came from the sequencing of a gene that complemented the so-called *act*III class of actinorhodin blocked mutants of *S. coelicolor* A3(2). The *act*III mutants represented one of two classes of mutants that were deduced to be interrupted in very early steps in the biosynthetic pathway because they failed to secrete any metabolite that could be converted by other mutants to actinorhodin, but they would convert to actinorhodin compounds secreted by four different classes of mutants, which were therefore deduced to be blocked at later steps.[64] Whereas the *act*I mutants were nonpigmented, and so were probably defective in assembly of the polyketide chain, the *act*III mutants produced a red diffusible pigment and were therefore likely to be defective in an early step in chain modification rather than in chain assembly itself. Sequencing of the

fragments of wild-type DNA into any available mutant blocked in a step of antibiotic biosynthesis, looking for complementation of the mutation, and then finding genes for the other steps of the pathway on the complementing fragments; and (2) cloning a library of DNA fragments from an antibiotic producer into a sensitive surrogate host (usually a derivative of *Streptomyces lividans* 66, which is a convenient, generally antibiotic-sensitive and easily manipulated strain), selecting resistant clones, and seeking biosynthetic genes linked to the resistance gene on the cloned DNA. By these procedures, the complete sets of biosynthetic genes for two further aromatic polyketides were isolated: the anthracycline, tetracenomycin (2) (*tcm*) from *Streptomyces glaucescens*,[30] and oxytetracycline (3) (*otc*) from *Streptomyces rimosus*.[31]

The availability of cloned DNA carrying the biosynthetic gene clusters for these three aromatic polyketides led to a test of the idea[29] that the sequences of different PKS genes, which are presumed to have diverged from a common ancestor (discussed in section VI), might be sufficiently conserved for a DNA fragment for one synthase to be used as a probe to isolate genes for others. The presumptive positions of the DNA encoding the polyketide KS and KR functions within the actinorhodin gene cluster had already been deduced[32] and so these DNA fragments (carrying the so-called *act*I and *act*III genes, respectively) could be used as hybridization probes against restriction digests of the *tcm* and *otc* cloned DNA.[33] Not only did strong cross-hybridization occur with the KS probe, but it recognized regions of the *tcm* and *otc* DNA clusters that had already been identified by complementation analysis as candidates for carrying the PKS genes. The KR probe hybridized to a second segment of the *otc* gene cluster, but not to any part of the cloned *tcm* DNA. This latter result was significant because tetracenomycin is one of the few polyketides that arise without any of the keto groups of the nascent carbon chain being reduced.

In the same study, the *act*I and *act*III probes were hybridized to Southern blots of restriction digests of total genomic DNA of 25 actinomycetes, including 18 polyketide producers and seven nonproducers: hybridizing bands were revealed in most, but not all, of the producers, and were absent from most, but not all, of the nonproducers. These results appeared to establish a strong, although imperfect, correlation between the presence in a particular actinomycete of DNA sequences that cross-hybridized with one or both of the *act* PKS probes and production of polyketides by that strain, suggesting that the probes could indeed be used to isolate further PKS genes. This was successfully accomplished for a second benzoisochromanequinone, granaticin (4), from *Streptomyces violaceoruber*.[33] Proof that the hybridizing sequences isolated from this strain indeed encoded the granaticin PKS was obtained by a gene disruption experiment. In this technique a fragment of a gene is introduced into the wild-type strain and a homologous crossover occurs between it and the resident copy of the same DNA to generate a nonproducing mutant (Figure 4). Similar experiments

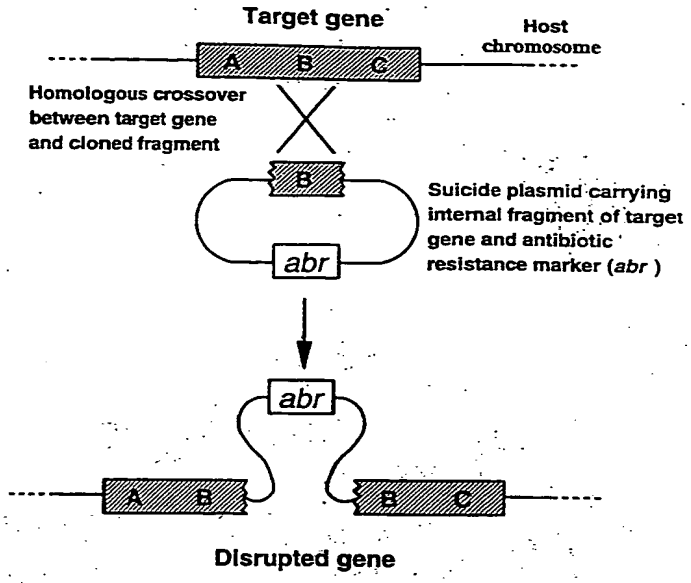


Figure 4. A simple gene disruption scheme. A plasmid carrying an internal fragment of a target gene to be disrupted, together with an antibiotic resistance marker gene, is introduced into an antibiotic producing host strain by transformation. The plasmid cannot replicate in the host (i.e., it is a "suicide" plasmid). When selection is made for the marker gene, by selecting for antibiotic resistance, only a recombinant in which the vector has become integrated by crossing-over, as shown, will be able to grow. (Figure kindly provided by M. J. Buttner.)

were also done with the producer of the macrolide milbemycin, *S. hygroscopicus* ssp. *aureolacrimosus*, from which DNA hybridizing to the *act* probes had also been isolated. In this case the results of the gene disruption experiment were less clear-cut, because the nonproducing phenotype was associated with unexpected chromosomal deletions from the region surrounding the presumptive PKS DNA of the wild-type host; in retrospect this result should perhaps have set off a warning bell.

Following publication of the Malpartida *et al.*[33] paper, the *act* probes were much in demand for attempts to isolate genes for further PKSs. Judging from the published literature, these attempts were all successful. However, it gradually became apparent that not all of the genes isolated using these probes actually encoded the PKS sought. For example, Arrowsmith *et al.*[34] isolated PKS genes from the producer of the polyether monensin, *Streptomyces cinnamonensis*, but their disruption failed to interfere with monensin biosynthesis. By now, the first example of a modular PKS, for biosynthesis of the macrolide erythromycin by *Saccharopolyspora erythraea*, had been discovered (see section III.A), and the *act* PKS probes failed to hybridize to the DNA encoding the erythromycin PKS; doubtless they did not hybridize to the monensin PKS genes either, but these have not been identified. Presumably genes for a PKS for an unidentified aromatic polyketide had been isolated from *S. cinnamonensis*. A second complication was revealed by the isolation and sequencing of DNA that complemented a mutation (called *whiE*) that prevents biosynthesis of the gray-brown spore pigment of *S. coelicolor* A3(2), resulting in a white spore phenotype.[35] The chemical nature

*act*III region revealed an open reading frame that would encode a protein resembling several known oxidoreductases, such as ribitol dehydrogenase from *Klebsiella aerogenes* and alcohol dehydrogenase from *Drosophila melanogaster*.[37] It was therefore deduced that the *act*III DNA encoded a discrete polyketide ketoreductase. By implication the PKS would have a type II structure with separate proteins for the individual reactions of chain assembly and modification.

Sequencing of the *act*I region, believed to encode the PKS KS, was delayed in relation to that of the corresponding PKS-encoding segments of the gene clusters for granaticin (*gra*) and tetracenomycin (*tcm*), which had meanwhile been isolated (see above). The sequences of the presumptive *gra* and *tcm* PKS-encoding DNA[45,42] immediately confirmed the type II nature of the PKSs by revealing open reading frames that would encode proteins resembling the condensing enzyme (the product of *fabB*) of the *E. coli* FAS[65] and the discrete acyl carrier proteins of the type II FASs of bacteria and plants. When the *act*I region was later sequenced, corresponding genes were identified.[38]

As well as putative structural genes for a KS, an ACP and (for *act* and *gra* but not *tcm*, as expected) a KR, a "mystery gene" was revealed lying immediately downstream of each KS gene and showing putative translational coupling with it (a situation in which there is overlap between the stop codon of the upstream gene and the start codon of the downstream gene: this arrangement is postulated to facilitate cotranslation of two bacterial genes to yield equimolar amounts of their protein products).[66] The downstream genes (initially called open reading frame 2: "ORF2") encoded proteins that resembled strongly the KSs ("ORF1"), but lacked the putative catalytic site for condensation (a characteristic motif based on a cysteine residue). At this stage, the idea that the "ORF2" sequences were nonfunctional pseudogenes was a formal possibility, but the subsequent finding that a mutation in *act*I-ORF2 abolished actinorhodin production (see section II.C)[67] ruled this out. I return to the role of "ORF2" later (section II.F.1).

A further function encoded by the *act* cluster, involved in cyclization of the nascent polyketide carbon chain, had been proposed by consideration of the phenotype of the so-called *act*VII mutants,[64] which were found to secrete an incorrectly cyclized shunt product, mutactin.[68] The sequence of the *act*VII gene,[38] and of its homologue in the *gra* cluster,[45] gave no clue to the biochemical role of the products of these genes. However, arguments described in section II.E, based on the structure of mutactin, suggested that these *act* and *gra* genes would encode a bifunctional cyclase/dehydratase. In contrast, a gene from the *tcm* cluster (*tcmN*), which would encode a protein whose N-terminal half resembled the corresponding regions of the *act* and *gra* cyclases/dehydratases, but whose C-terminal half resembled instead O-methyltransferases, would encode a bifunctional cyclase/O-methyltransferase.[43,69] I return later to the role of the *act*VII gene also.

In describing the architecture of the *act* cluster, a sixth gene also needs to be introduced. This is *act*IV, whose function as a cyclase became apparent only as a result of the genetic engineering studies that are discussed below. The arrangement of the six *act* PKS genes, and for comparison those of the genes that encode other actinomycete aromatic PKSs, are shown in Figure 5. There are striking resemblances in overall architecture between the various clusters, but also differences in the arrangements of some of the homologous genes.

C. Would "Hybrid" Synthases Work?

The potential for producing new, "hybrid" antibiotic structures by engineering novel combinations of antibiotic biosynthetic genes from different organisms was first demonstrated when segments of the *act* biosynthetic gene cluster, or the whole cluster, were transferred into the streptomycetes that produce medermycin or granaticin (and dihydrogranaticin).[70,71] The hybrid compounds discovered in these studies, and in later experiments involving the carbomycin and spiramycin biosynthetic genes,[72] were polyketides, but the genes that were recombined to make them encoded late, tailoring steps in the biosynthetic pathways (catalyzed by individual enzymes such as reductases, hydroxylases, or other group transferases) rather than steps in polyketide chain assembly and immediate postassembly modification. While there was hope that more radical engineering of product structure could be achieved by bringing together subunits of type II PKSs from different organisms to form functional hybrid synthases, this was not a foregone conclusion; specific protein–protein interactions might have evolved over long periods of evolutionary history to ensure that a set of subunits could work efficiently together to form a functional synthase. The first indication that this would not necessarily be a barrier to effective "mixing-and-matching" of PKS subunits was provided by the complementation of a mutation in the *act*III (KR) gene of *S. coelicolor* by the homologous DNA from the granaticin gene cluster and what was at that time presumed to be the milbemycin gene cluster to produce pigments similar or identical to actinorhodin.[33] A reciprocal experiment was later reported by Bartel *et al.*,[73] who transformed a KR-negative mutant of *Streptomyces galilaeus* that produced 2-hydroxyaklavinone with the *act*III gene and restored aklavinone production. Importantly, they went further[73] in demonstrating the production of a novel (for *S. galilaeus*) polyketide, aloesaponarin II (see below) by wild-type *S. galilaeus* transformed with a plasmid carrying *act*I-ORF1 and *act*I-ORF2;[38] this was an example of "mixing-and-matching" in which there must have been effective cooperation of the *act* KS with heterologous subunits.

In order to investigate further the possibilities for heterologous PKS subunit interactions, attempts were made to complement mutations in each of the five so far identified *act* PKS subunit genes (*act*III, *act*I-ORF1, *act*I-ORF2, *act*I-ORF3, and *act*VII) by the corresponding genes from the *gra* set of *S. violaceoruber*. Since the assembly, ketoreduction, and cyclization of the polyketide chains that are later

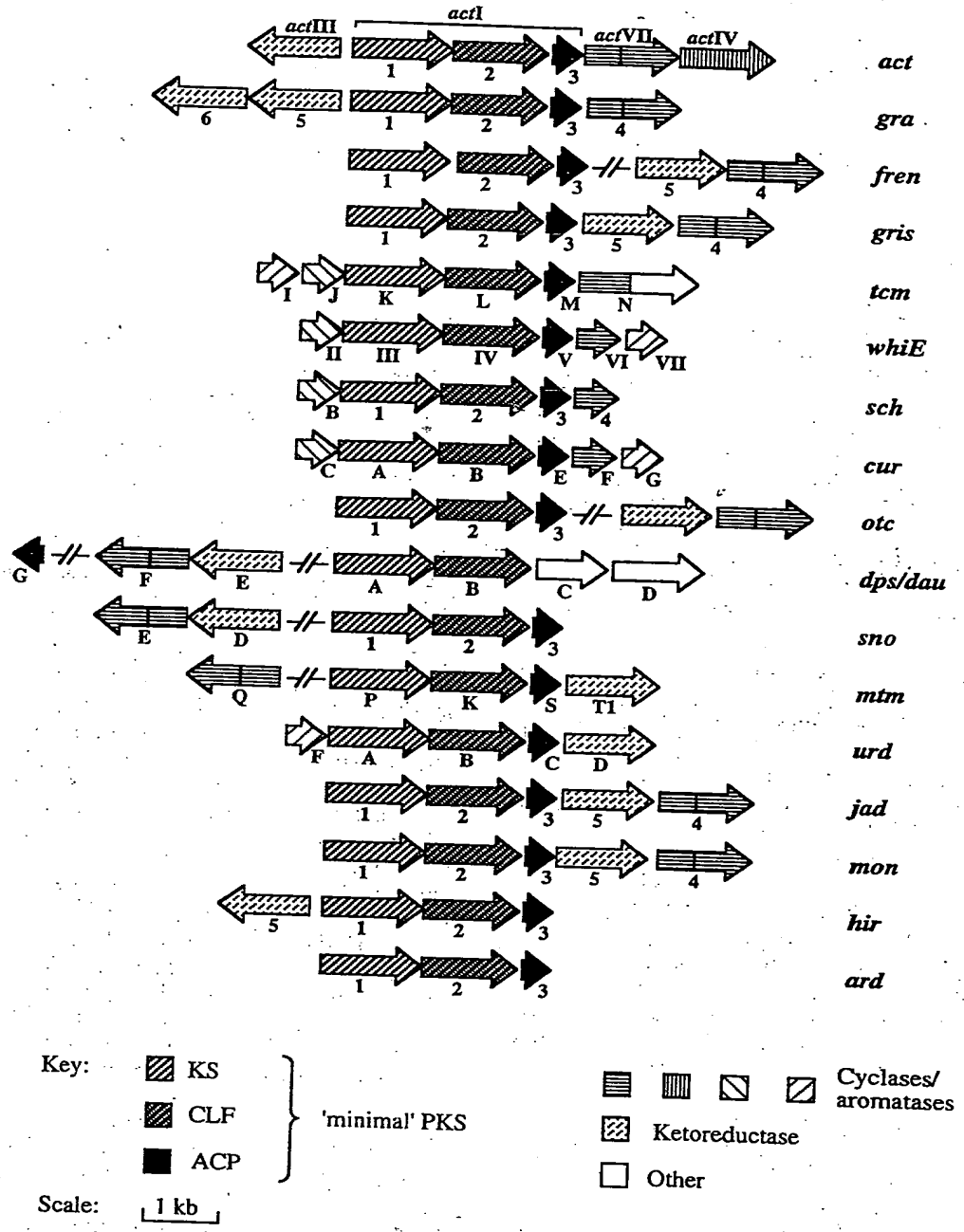


Figure 5. Organisation of gene clusters for type II PKSs in actinomycetes (for references and nucleotide sequence accession numbers see Table 1). In general, only those genes known to be involved in carbon chain assembly, reduction, aromatization, or cyclization are shown. Exceptions are the genes or domains labeled "other": the C-terminal half of *tcmN* encodes an O-methyltransferase, and *dpsC/D*, referred to in the text, play unknown roles. *gra*-ORF6, which resembles *gra*-ORF5 in overall sequence but lacks a nucleotide-binding motif characteristic of ketoreductases,[45] might cooperate with the ORF5 protein in its KR function.[67] Note that *dpsA* and *dpsB* may be translationally coupled, like most of the other KS/CLF gene pairs, because they are separated by only 3 bp,[47] while *dauA* and *dauB* are separated by about 100 bp,[53] and so are uncoupled, like their *fren* homologues.[50] (Figure kindly drawn by W. P. Revill.)

tailored to either actinorhodin or granaticin were expected to be identical,[74,75] these experiments would test for the productive interaction of heterologous PKS subunits without requiring the hybrid PKS to generate a novel product, which could perhaps have failed for purely chemical reasons. The results were very encouraging.[67] Complementation of the *act*III (KR) mutant by the *gra* KR gene (ORF5: Figure 5) was confirmed, and complementation of *act*VII (cyclase/dehydratase) mutants by the corresponding *gra* gene (ORF4) was also clearly demonstrated. The situation was more complex for *act*I, because the 13 available mutants had not previously been assigned to one or other of the three ORFs. Complementation tests using *gra*-ORF1, *gra*-ORF2, and *gra*-ORF3 allowed a number of conclusions to be drawn: (1) several *act*I-ORF1 mutants were identified by heterologous complementation by *gra*-ORF1; (2) a single *act*I-ORF2 mutant was clearly identified by complementation by *gra*-ORF2, thereby demonstrating not

only heterologous complementation but also the essential nature of the ORF2 gene; and (3) none of the *act*I mutants was complemented by *gra*-ORF3, indicating either that the set of *act*I mutants included no examples of lesions in this small ACP gene, or a failure of heterologous complementation.

These early results were soon extended by (1) unambiguous demonstration of the requirement of the *act*I-ORF3 gene product, the ACP, in a functional PKS;[76] (2) functional replacement of the *act* ACP by the corresponding ACPs from the *gra*, *tcm*, *otc*, and putative *fren* PKSs[77] and even, at a low level, by a putative FAS ACP from *Saccharopolyspora erythraea*;[76] and (3) complementation of an *act*I-ORF1 (KS) mutant, not only by the corresponding *gra* gene as reported earlier, but by homologues from the *otc* and *whiE* (spore pigment) gene clusters.[78] The stage was now set for a systematic mix-and-match approach in which the structures of polyketides generated by recombinants containing hybrid PKS gene clusters could begin to reveal the programming rules for the type II PKSs,[79,80] but first a new genetic test system was needed.

D. A Special Host–Vector System for Construction and Expression of Recombinant PKSs

The studies just reviewed, which established the feasibility of a systematic search for novel polyketides generated by hybrid PKSs, were carried out in strains of *S. coelicolor* with point mutations or deletions in individual PKS genes of the *act* cluster, which consists of 22 structural, regulatory, export, and resistance genes.[37,38,81–83] It soon became apparent that these strains were far from ideal for chemical analysis, since recombinants could potentially generate complex mixtures of end products, shunt products, and degradation products through the action of the *act* genes that were still functional in the mutants.[76] It was therefore decided to build sets of aromatic PKS subunits and to express them in the absence of the enzymes for tailoring pathway steps, whose natural role is to diversify the primary products of the PKSs themselves. A special host–vector system was engineered for this objective. The host is an *S. coelicolor* A3(2) derivative (strain CH999) from which the entire set of 22 *act* genes (except for one, *act*VI-ORFA, at the extreme lefthand end of the cluster)[82] has been deleted and replaced by a convenient marker gene for erythromycin resistance.[84] (This was achieved by a variation in the gene disruption procedure in which a segment of DNA from each end of the cluster is cloned on a suicide vector on either side of the marker gene to form a replacement "cassette". Double crossing-over between this cassette and the corresponding chromosomal sequences gives rise to the desired replacement.) Into this host is introduced, by a standard protoplast transformation procedure,[85] any of a series of plasmids carrying the desired sets of PKS subunit genes. These plasmids are based on the replicon of a stably inherited, low copy number *S. coelicolor* plasmid, SCP2*,[86] onto which the *tsr* marker for thiostrepton resistance in *Streptomyces* is cloned. (Low copy number is often preferable to multicopy cloning of antibiotic biosynthetic genes in *Streptomyces*, probably to avoid clone instability caused by physiological stress.) The vector also carries an *E. coli* origin of replication from ColE1, and the ampicillin resistance gene (*bla*) for selection in *E. coli*; thus rapid genetic engineering can be carried out in *E. coli*, before transfer of the finished constructs to *S. coelicolor* CH999. A further component of the plasmid is the *act*II-ORF4 gene, the natural pathway-specific activator of the *act* biosynthetic genes;[81] its product serves to activate transcription, in an appropriate, developmentally controlled manner, from the *act*I and *act*III promoters, which form a divergent pair.[37,38] Downstream of one or both of these promoters are cloned the desired PKS subunit genes. In pRM5 (Figure 6), the founding member of the family of plasmids,[84] this set consists of the *act*III, *act*I (3 ORFs), *act*VII and *act*IV genes; they are cloned in their natural arrangement (Figure 5), except for the translational decoupling of some pairs of genes and the introduction, between each pair, of convenient unique restriction sites to aid the construction of recombinants by the exchange of genes. CH999 carrying pRM5 or a series of plasmids based on it has generated a great deal of new information about the roles of the aromatic PKS subunits, which will now be summarized. Although not following precisely the historical sequence, it will be convenient first to describe experiments with recombinants carrying subsets of PKS genes from the same strain, and then experiments with hybrid clusters containing genes originating from two or more *Streptomyces* species.

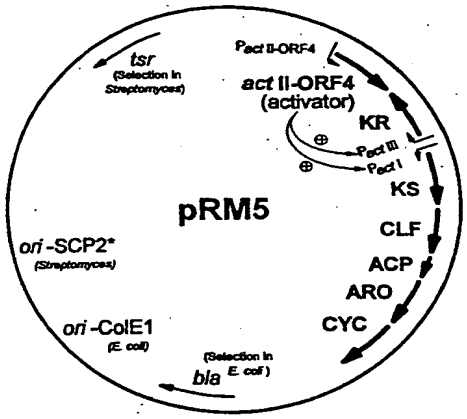


Figure 6. The vector pRM5.[84] The plasmid is bifunctional for replication in both *E. coli* (ColE1 origin of replication (*ori*) and β-lactamase (*bla*) gene for selection) and *Streptomyces* (SCP2* origin of replication and thiostrepton resistance gene (*tsr*) for selection). The cloned PKS genes to be expressed in *Streptomyces*—in this example the *act* KS, CLF, ACP, ARO, CYC, KR genes—are cloned downstream of their natural promoters, P_{actI} and P_{actIII}, which are transcriptionally activated by the product of the *act*II-ORF4 gene. (Figure kindly drawn by T. Kieser.)

E. Recombinants Carrying Nonhybrid Subsets of PKS Genes

CH999 carrying pRM5 was found to produce a compound, aloesaponarin II, which had earlier been identified by Bartel *et al.*[73] in the original *act*VI mutant class. Cosynthetic studies had suggested that the *act*VI mutants were blocked immediately

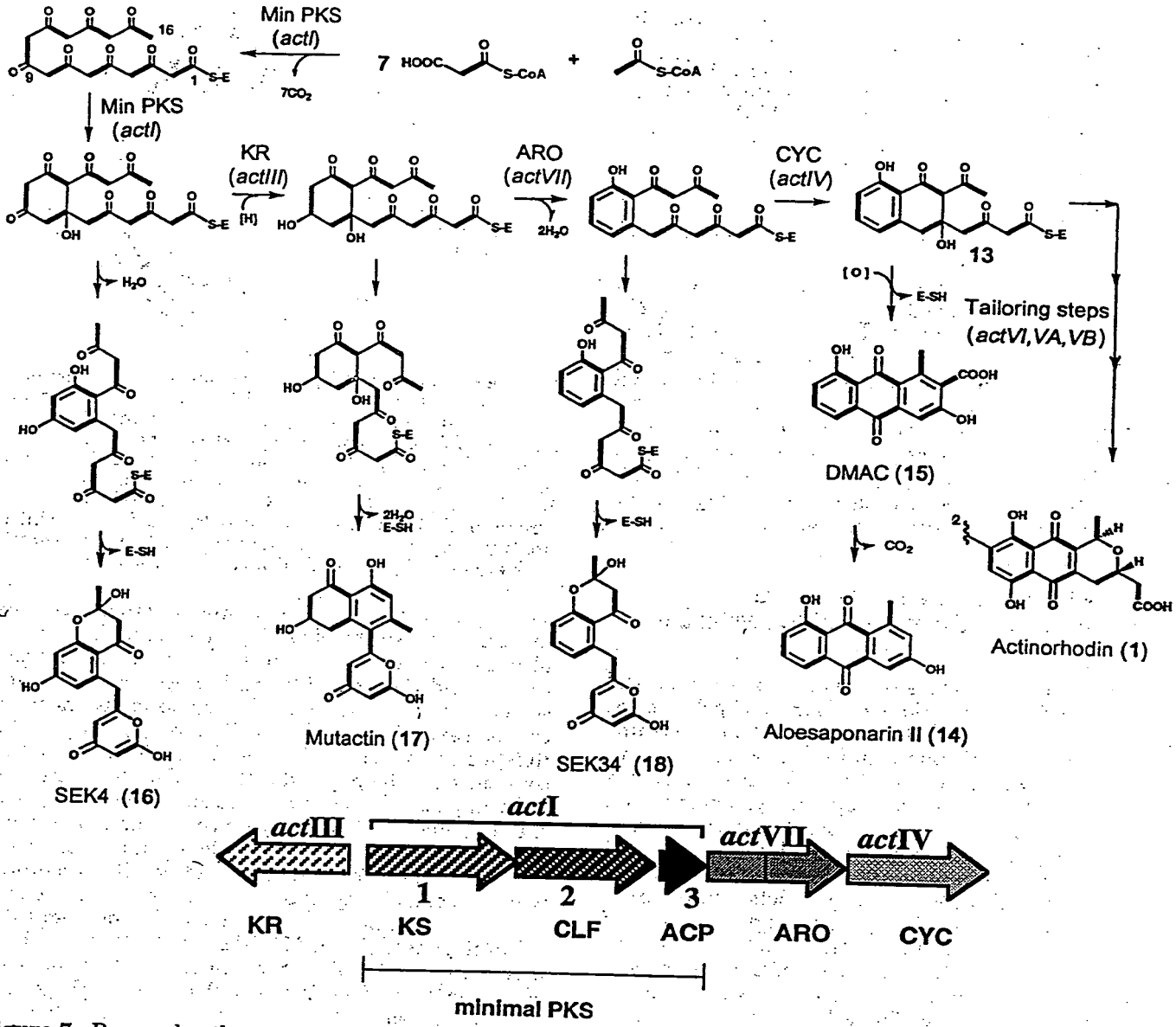


Figure 7. Proposed pathway for biosynthesis of the polyketide intermediate **13** of actinorhodin (**1**) by the PKS encoded by the *act* genes of *S. coelicolor* A3(2): *act*I-ORF1, *act*I-ORF2, *act*I-ORF3, *act*III, *act*VII, *act*IV. One starter acetyl CoA and seven chain extender malonyl CoA molecules are thought to be assembled into an unreduced polyketide carbon chain by the minimal PKS (min PKS, consisting of the ketosynthase (KS), chain-length factor (CLF), and acyl carrier protein (ACP)). The minimal PKS is also presumed to catalyze formation of the first carbocyclic ring. The products of *act*III (the ketoreductase for C-9, KR), *act*VII (the aromatase for the first ring, ARO), and *act*IV (the cyclase for the second ring, CYC) are proposed to carry out the reactions shown to generate **13**. The hypothetical pathway is deduced from the structures of the shunt products, **14**–**18**, produced by appropriate mutant or recombinant strains (see text). (Adapted from ref 87.)

after the *act*IV mutants in the biosynthetic sequence: *act*I → III → VII → IV → VI → V → actinorhodin.[64] Aloesaponarin II was postulated to arise as a shunt compound from the hypothetical pathway intermediate (**13**: Figure 7) produced by the enzymes encoded by the *act*I + III + VII + IV genes.[73] This intermediate would be an octaketide (i.e., C_{16}) that had undergone the correct C-9 reduction and dehydration and correct formation of the two carbocyclic rings characteristic of the actinorhodin half-molecule. In the *act*VI mutants, **13** was postulated to undergo spontaneous formation of a third carbocyclic ring to generate an anthraquinone system, followed by a decarboxylation, to yield aloesaponarin II (**14**).[73] The isolation of aloesaponarin II in significant quantities (up to 100 mg/L) from cultures of CH999/pRM5[84] was not only an encouraging start for the project but was entirely consistent with these earlier findings because pRM5 in fact carries precisely the set of genes (*act*I + III + VII + IV) that were postulated to be responsible for formation of aloesaponarin II in the *act*VI-blocked mutants. Moreover CH999/pRM5 also made 3,8-dihydroxy-1-methylanthraquinone-2-carboxylic acid (DMAC, **15**), the expected undecarboxylated precursor of aloesaponarin II.[84] These results established pRM5 as the ideal starting point for the recombinant construction project, because it evidently contained all the genes needed to encode a correctly programmed PKS without the complication of the tailoring enzymes.

Among the set of genes present on pRM5, it turned out that all three of the genes in the *act*I segment were needed to form any recognized product in the CH999 host; they were therefore called the "minimal PKS" genes.[88] (It was a happy chance that the "classical" *act*I mutants, although potentially representing lesions in three different genes—ORF1, ORF2, and ORF3—all correspond to mutations in the minimal PKS, thereby providing a retrospective rationalization for designating them *act*I-ORF1, *act*I-ORF2, and *act*I-ORF3; and luckily the other relevant mutant classes—*act*III, *act*VII, and *act*IV—each represent mutations in a single gene.) The product of the *act* minimal PKS was SEK4 (**16**),[89] with an unreduced (as expected) C_{16} chain and just the first carbocyclic ring correctly formed by aldol condensation between C-7 and C-12; the rest of the structure, as in all the shunt products to be described, would have arisen by the most probable uncatalyzed chemistry, in this case to give a hemiketal at the methyl end of the chain and a pyrone at the carboxyl end.

When the *act*III gene was added to the minimal PKS gene set, its predicted product, the KR, reduced the C-9 carbonyl, giving the shunt product mutactin (**17**).[87] This also agreed with precedent, because the construct carrying *act*I + III would be equivalent to an *act*VII mutant, in which mutactin had originally been identified.[68] Mutactin is bicyclic, but the second ring is not the one found in actinorhodin, leading Zhang *et al.*[68] to postulate that the *act*VII gene encodes a second ring cyclase; C-9 is also not dehydrated, so the idea of a bifunctional cyclase/dehydratase was born.[69]

When McDaniel *et al.*[87] added *act*VII to *act*I + III they identified a further novel shunt product, SEK34 (**18**), which still lacked the correct second ring, but which, unlike mutactin, now had an aromatic first ring; the *act*VII protein was therefore renamed an "aromatase", which would do its job by carrying out *two* dehydrations on the first ring. This proposal was speculatively linked to the finding that the *act*VII protein may have two separate functional domains representing the N- and C-terminal halves. Their primary sequences are similar, suggesting that they might have originated by an ancestral gene duplication, a feature that was noticed first for the corresponding gene of the frenolicin (*fren*) PKS cluster, but is retrospectively apparent also in its homologues in the other gene clusters.[50] I return to this point later.

Finally, when the *act*IV gene was added to the growing set of *act* genes (I + III + VII), as in pRM5 itself, the production of DMAC and aloesaponarin II implied that *act*IV encoded the true second ring cyclase.[87] The *act*IV protein had earlier been found to have sequence similarities with some Zn^{2+}-containing β-lactamases,[38] and it was pointed out that the second ring cyclization involves an aldol condensation and that a class of aldolases are Zn^{2+}-containing enzymes![87]

That the *S. coelicolor* CH999 host was not providing a unique background for such experiments had been indicated by the results of Bartel *et al.*[73] and was reinforced by those of Kim *et al.*;[78] both groups identified mutactin and aloesaponarin II as the products of suitable combinations of the *act* PKS subunits in a heterologous, polyketide nonproducing host, *Streptomyces parvulus*. This system was used[78] to confirm that the presumed active site Cys in the *act*I-ORF1 (KS) protein was essential for polyketide synthesis, as expected from sequence comparisons with other condensing enzymes (a result also found for the corresponding residue in the *tcm* KS);[90] and that the presence of the *act*I-ORF2 gene was also required for any product formation.

In other experiments, subsets of the *tcm* PKS gene cluster were expressed in the CH999 host, leading to some comparable conclusions for their roles to those of the *act* genes. For example, the *tcm* homologues of the three *act*I gene products yielded an unreduced C_{20} compound, SEK15 (**19**: Figure 8), in contrast to the C_{16} SEK4 (**16**) made by the corresponding *act* gene products.[89] Clearly chain length was being determined by the minimal PKS, but it needed some of the first mix-and-match experiments to try to attribute this property to a specific subunit of the PKS. Similarly, the finding of the natural first ring in SEK4, with a C-7/C-12 condensation, had suggested that the minimal PKS could control first ring formation, but SEK15 also arose by C-7/C-12 first ring condensation, even though the natural course of tetracenomycin biosynthesis in *S. glaucescens* involves a C-9/C-14 condensation to produce the first ring. Again, mix-and-match experiments were needed to throw light on control of the regiospecificity of cyclization. These will now be described.

F. Recombinants Carrying Hybrid Sets of PKS Genes

The first hybrids to be constructed using the CH999/pRM5 host—vector system involved all possible combinations of the *act*I-ORF1, -ORF2, and -ORF3 genes and their homologues from the *gra*, *tcm*, and *fren* PKS gene sets.[84,91] The *act*III (KR) gene was present in all these early constructs, as were the *act*VII and *act*IV genes (later identified as aromatase and cyclase genes, respectively, as just described). The results were dramatic, in that the majority of the recombinants generated polyketide products; these included the known octaketides DMAC and aloesaponarin II, and three novel compounds: a decaketide (RM20: **20**), a nonaketide (RM18: **21**), and an octaketide (RM18b: **22**). Over the last three years more than 30 novel compounds have been generated by the mix-and-match approach. These have not only established combinatorial biosynthesis as a route to interesting new compounds ("unnatural natural products")[99,100] but have thrown considerable light on the roles of the various PKS subunits in programming.

1. Chain-Length Determination: Identification of the Chain-Length Factor

The most striking conclusions to be drawn from the first set of results concerned the determination of carbon chain length. One deduction was very clear:

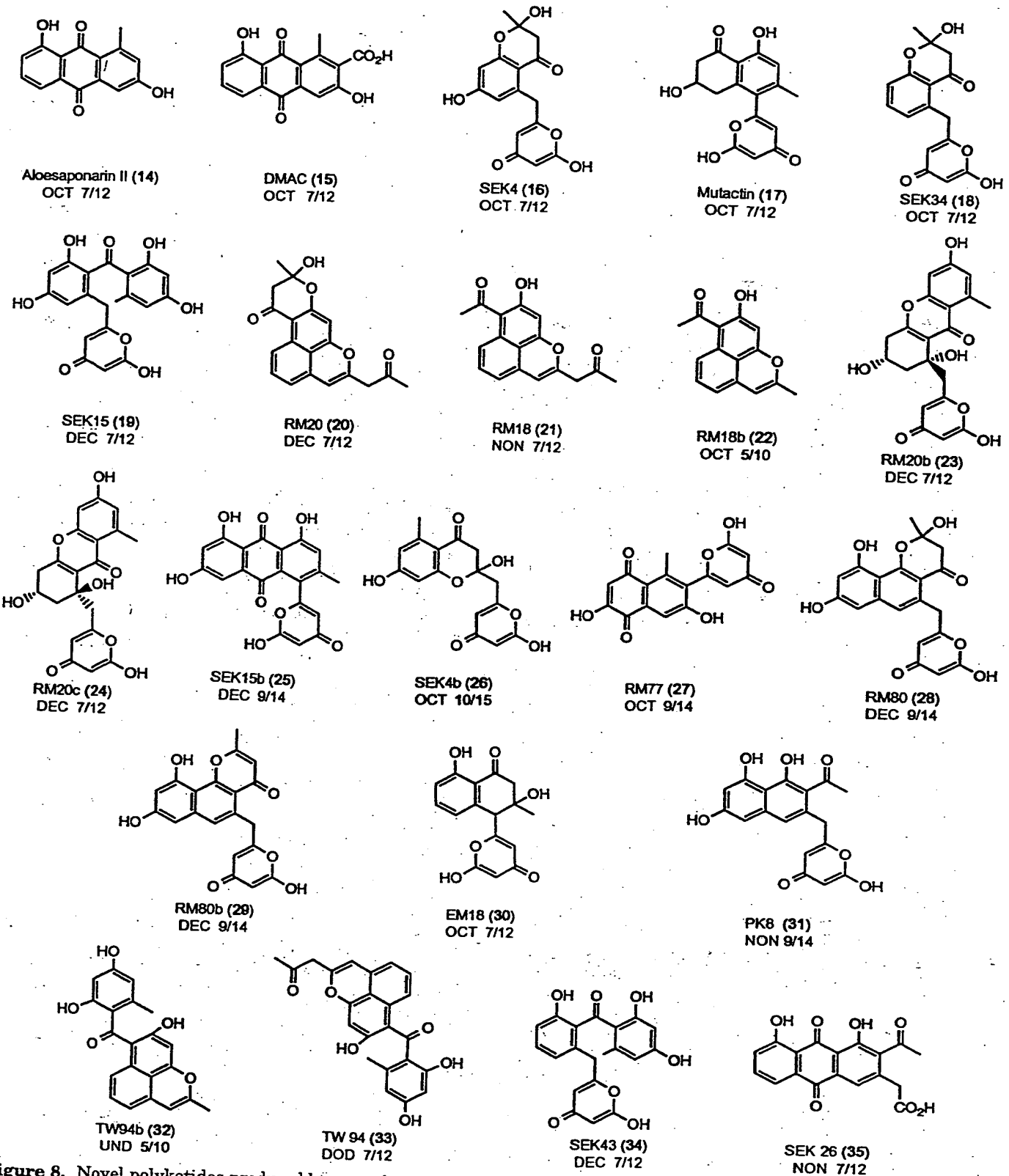


Figure 8. Novel polyketides produced by recombinant *S. coelicolor* CH999 strains containing plasmids related to pRM5 and carrying genes encoding various combinations of type II PKS subunits from the same or different *Streptomyces* species. Under the name of each metabolite is given the number of acetyl residues needed to build the carbon chain (OCT, octaketide; NON, nonaketide; DEC, decaketide; UND, undecaketide; DOD, dodecaketide), and the carbon atoms (counting from the carboxyl end) between which the first ring-closure occurs. References: **14**;[73,84] **15** and **20**;[84] **16** and **19**;[89] **17**;[68] **18**;[87] **21** and **22**;[91] **23** and **24**;[92] **25**;[88] **26**;[93] **27**–**29**;[94] **30**;[95] **31**;[96] **32** and **33**;[97] **34** and **35**.[98]

the source of the acyl carrier protein (the product of *act*I-ORF3 and its homologues) did not affect chain length, which always followed the source of ORF1 and ORF2 when these were from the same source:

the chain length was C_{16} for *act* and *gra*, C_{20} for *tcm* and a mixture of C_{18} and C_{16} for *fren*. This last finding of both nona- and octaketides (RM18 (21) and RM18b (22)) for the *fren* PKS was significant because *S. roseofulvus*, the natural host for the genes, produces a mixture of the nonaketide frenolicin and the octaketide nanaomycin. (The characterization of RM18 and RM18b in the recombinants actually provided evidence that the *fren* genes did in fact encode a PKS responsible for biosynthesis of both frenolicin and nanaomycin, a problem that could not be resolved by gene disruption because of the intractability of *S. roseofulvus* to genetic transformation.)[50] Construction of heterologous ORF1 and ORF2 combinations met with mixed success. Both combinations of *act* and *gra* were functional, but yielded no information on chain-length determination because both the natural PKSs produce octaketides. The combinations of *fren* or *tcm* ORF1 with *act* ORF2 also yielded products. In contrast, the reciprocal combinations, *act* ORF1 with *fren* or *tcm* ORF2, were inactive, as were both combinations of ORF1 and ORF2 from the *tcm* and *fren* clusters. Nevertheless it was possible to conclude that the ORF2-encoded PKS components determined chain length, *"at least in part"*.[84,91] This was because the combination of *act* ORF1+ORF2 produced an octaketide; this was not changed by substituting *act* ORF1 by either *fren* or *tcm* ORF1; but when both *act* ORF1 and *act* ORF2 were replaced by the corresponding *fren* or *tcm* genes, a mixture of a nonaketide and an octaketide, or a decaketide, respectively, was produced. The product of the ORF2 gene was therefore named the "chain-length factor" (CLF) to provide a convenient epithet for this hitherto "mystery" gene.

It was originally suggested[42] that the ORF1- and ORF2-encoded PKS subunits might associate to form a heterodimeric KS, reminiscent of the known homodimeric *E. coli* enzymes,[101] and this remains a reasonable hypothesis. The finding that these two proteins are more "choosy" in their ability to form an active recombinant PKS than other components such as the ACP is consistent with this idea, since specific protein–protein interactions could well be required to assemble a functional dimer from these similar but not identical subunits. It may be relevant that, while the arrangement of the genes in the different PKS clusters varies (Figure 5), there is so far no exception out of at least 18 examples to the finding of the CLF gene just downstream of the KS gene, and coupled to it in all except the *fren* and *dau* clusters. Phylogenetic analysis clearly implies that the ORF1 and ORF2 genes originated by a gene duplication in an ancestral PKS gene set and have then diverged (Figure 9). It also appears that the ORF2 proteins are more diverged from each other, and from the presumed common ancestor shared with the ORF1 proteins, than are the ORF1 proteins from each other.[51] This is consistent with, although not of course proving, the idea that the ORF2 proteins play a more specific role in chain-length determination than the ORF1 proteins, which would serve primarily to catalyze the condensation reaction itself.

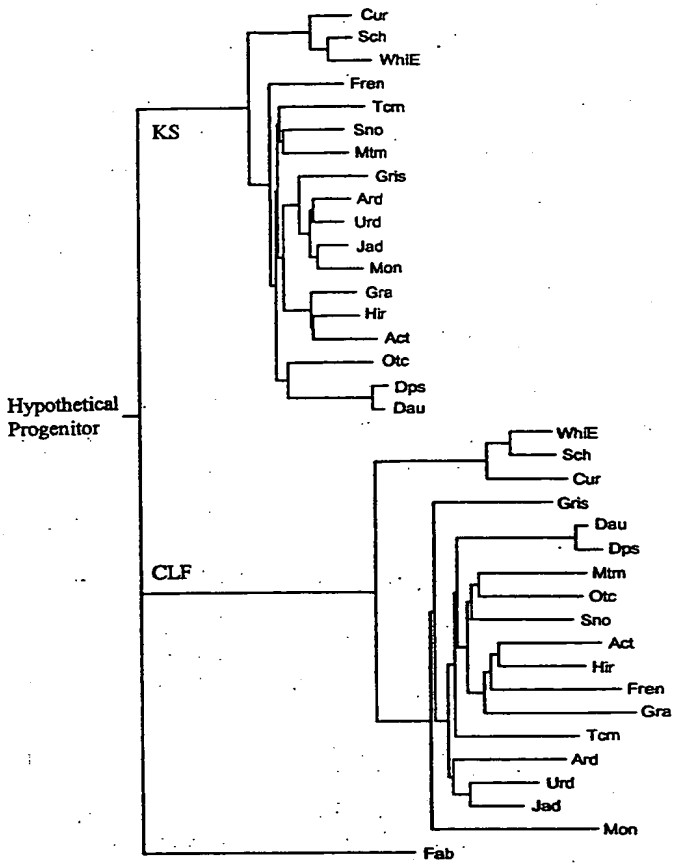


Figure 9. Phylogenetic tree of amino acid sequences of KS and CLF subunits of actinomycete type II PKSs. For sources of data, see Table 1. The tree was constructed using the PHYLUP 3.5 package with the "Dayhoff PAM distance matrix" and the Fitch–Margoliash method for tree construction, and the *S. glaucescens* putative FAS condensing enzyme (Fab)[101a] as outlier. (The number of amino acid substitutions is proportional to the lengths of the horizontal lines; the lengths of the vertical lines are arbitrary.) (Figure kindly provided by W. P. Revill.)

2. Role of the Minimal PKS in First Ring Cyclization

Establishing the correct fold of the nascent carbon chain to close the first carbocyclic ring of the aromatic polyketides has long been recognized as a key step in controlling product structure.[102,103] The finding of SEK4, with its C-7/C-12 ring closure, as the product of the *act* minimal PKS suggested that these three proteins are capable of establishing the correct first ring cyclization. This conclusion was later modulated following isolation of other novel metabolites made by minimal PKSs.

The product of the *tcm* minimal PKS plus the *act* KR was initially recognized as RM20 (20),[84] and later two further products, RM20b (23) and RM20c (24), were identified.[92] All three showed C-7/C-12 first ring closure. When the *tcm* minimal PKS was expressed without the *act* KR (but with the *act*VII and *act*IV genes) SEK15 (19), also with C-7/C-12 closure, had been found[89] (see section II.E), but small quantities of SEK15b (25), with C-9/C-14 closure, also occurred; and when the *tcm* minimal PKS was expressed alone, the ratio of SEK15 and SEK15b became approximately 1:1.[88] In contrast, in its natural host, *S. glaucescens*, the complete *tcm*

PKS—but in the absence of the unnatural *act*KR, ARO, and CYC—produces only the C-9/C-14 closure characteristic of tetracenomycin (**2**: Figure 3).

These findings gave rise to the idea that a minimal PKS can control first ring cyclization, but that this control can be modulated by other PKS subunits, which need not necessarily act as enzymes on the nascent carbon chain (for example in the case of the *act*VII and *act*IV proteins just mentioned), but which may modify the interactions between the subunits of the PKS complex and its interactions with the growing chain. Another example of the same phenomenon was postulated following the discovery of a further metabolite, SEK4b (**26**), produced in equal quantities along with the previously identified SEK4 by the *act* minimal PKS.[93] In this molecule the normal C-7/C-12 cyclization is replaced by an unusual C-10/C-15 first cyclization of the methyl end of the carbon chain. Production of SEK4b was ascribed to uncatalyzed cyclization of this end after its premature release from the PKS. Again, the outcome was influenced by other PKS subunits: the *act* minimal PKS alone produced a ratio of 1:5 for SEK4 and SEK4b, but in the presence of the *act*VII and *act*IV proteins, the proportion of the naturally folded SEK4 was much greater.[93]

3. Aromatases and Cyclases

As described above (section II.E), studies of subsets of the *act* PKS genes alone led to characterization of the *act*VII gene product as an aromatase for the first ring.[87] These studies also indicated that the *act*IV gene product is a second ring cyclase, catalyzing an aldol condensation between C-5 and C-14 to yield the bicyclic fused ring structure of benzoisochromanequinones like actinorhodin. Since then, PKS subunits from other streptomycetes have been characterized as aromatases and/or cyclases.

Several PKS gene sets, besides the *act* set, contain homologues of *act*VII (Figure 5). These PKSs synthesize the C-9-reduced octa-, nona-, and decaketide carbon chains of the benzoisochromanequinones (actinorhodin, granaticin, frenolicin, and griseusin), as well as the decaketides of oxytetracycline, anthracyclines, jadomycin, mithramycin, and the unknown aromatic polyketide of *S. cinnamonensis*. They tend to show the resemblance between the N- and C-terminal halves of the proteins, first recognized in the *fren* example and postulated to have arisen by an ancestral gene duplication (Figure 10).[50] They can be described as "didomain" aromatases,[104] presumptively acting by the abstraction of two molecules of water. Striking support for the existence of two separate domains was recently obtained by artificially expressing them as separate proteins, which functioned together *in vivo*.[104] In contrast, the *tcm* cluster contains a gene, *tcmN*, whose product has an N-terminal half homologous with that of the *act*VII protein, fused to a C-terminal O-methyltransferase for a tailoring step, but the N-terminal half is capable of functioning by itself as a "monodomain" aromatase.[43] A homologous monodomain protein is encoded by the *whiE*-ORFVI gene and its homologues in the *sch* and *cur* clusters (Figures 5 and 10) which are components of PKSs that probably, like the *tcm*

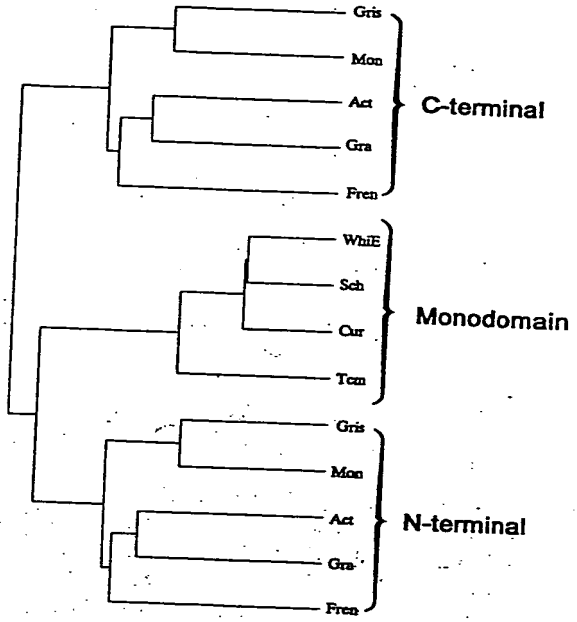


Figure 10. Dendrogram of the deduced amino acid sequences of the N-terminal and C-terminal halves of the gene products of some presumed didomain aromatase subunits of actinomycete type II PKSs; the N-terminal half of the tetracenomycin aromatase/O-methyltransferase; and the whole sequences of the presumed monodomain aromatase subunits of the WhiE, Sch, and Cur PKSs. For sources of data, see Table 1. (Reprinted with permission from ref 50. Copyright 1994 American Society for Microbiology.)

PKS, synthesise unreduced polyketide chains, whose first rings would aromatize spontaneously.

Both *tcmN* and *whiE*-ORFVI have been combined with different minimal PKS gene sets (*act*, *fren*, and *tcm*), with and without the *act* KR, and various novel polyketide products were identified. From their structures it was deduced that the *tcmN* protein can influence the regiospecificity of first ring closure of unreduced but not reduced carbon chains: exclusively C-9/C-14 closure was observed (to give RM77: **27**) when *tcmN* was added to the *act* minimal PKS gene set,[94] instead of SEK4 (C-7/C-9) in its absence, providing a further example of the "modulation" phenomenon. However, the primary function for the *tcmN* protein was deduced, from the structures of RM80 (**28**) and RM80b (**29**) (both products of the *tcm* minimal PKS together with *tcmN*), to be as a *second ring* aromatase.[94] The homologous *whiE*-ORFVI protein also influenced the regiospecificity of first ring cyclization of unreduced but not reduced carbon chains, and interestingly it could catalyze first ring aromatization of a reduced chain (which it presumably does not normally see) to generate the novel polyketide EM18 (**30**).[95]

4. Flexibility of Chain-Length Determination

A striking feature of most PKSs is the fidelity with which they control the carbon chain length of their products. Two interesting exceptions have emerged from the construction of recombinants in the *S. coelicolor* host–vector system. The first concerns the *fren* minimal PKS which, as described above, was found to construct both octa- and nonaketides when

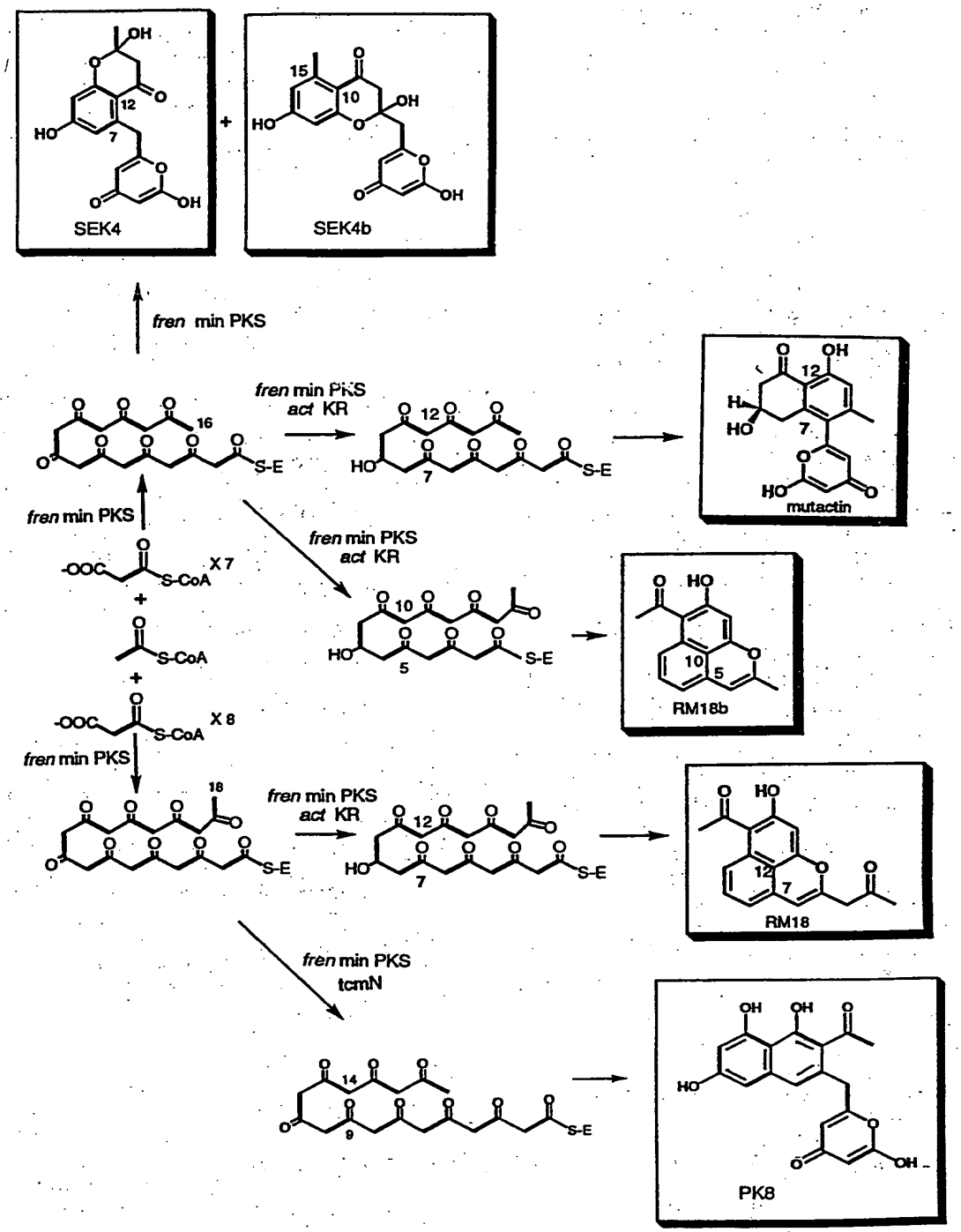


Figure 11. Carbon chain assembly and first ring cyclization patterns that lead to six novel polyketides catalyzed by the *fren* minimal PKS, alone or modulated by the presence of the *act* KR or *tcmN* aromatase.[96] Note that the *fren* minimal PKS alone builds only octaketides (with C-7/C-12 or aberrant C-10/C-15 ring closure), SEK4 and SEK4b; in the presence of the *act* KR it builds both octaketides (with either C-7/C-12 or C-5/C-10 ring closure), mutactin and RM18b, and a nonaketide (with C-7/C-12 ring closure), RM18; and in the presence of the *tcmN* aromatase it builds only a nonaketide (with C-9/C-14 ring closure), PK8. See text for details.

heterologously expressed along with the *act* KR, as it is presumed to do along with the *fren* KR in its natural *S. roseofulvus* host.[91] Later it was discovered that the choice of chain length could be constrained by other PKS subunits, in either direction (Figure 11): the *fren* minimal PKS *alone* produced only the octaketides SEK4 and SEK4b, whereas addition of *tcmN* caused the (almost) exclusive production of a novel nonaketide, PK8 (**31**).[96]

The second example involves the *whiE* minimal PKS which, when cloned in the CH999 expression system, was found to be capable of synthesizing both C_{22} undeca- and C_{24} dodecaketides, TW94b (**32**) and TW94 (**33**).[97] In this case the chain length could be constrained to be exclusively C_{24} by addition of the *whiE*-ORFVI gene discussed above in connection with its role in influencing first ring cyclization and as a first ring aromatase.[105] This might be taken to imply

that the normal spore pigment of *S. coelicolor*, the product of the complete *whiE* gene set, is a dodecaketide.

5. Choice of Starter Unit

The starter unit for the majority of aromatic polyketides is thought to be acetate, but there are exceptions: the best documented is the use of propionate for anthracyclines like daunorubicin and doxorubicin made by strains such as *S.* sp. C5 and *S. peucetius*. In other examples the situation is less clear: in *S. rimosus* the starter unit for oxytetracycline biosynthesis by the *otc* PKS is commonly believed to be malonamate, but might be malonate;[106] and for frenolicin (**6**, Figure 3), in *S. roseofulvus*, with its fully reduced methylene group at C-17, there has been speculation that the starter unit might be butyrate, obviating a need for a full cycle of reduction at this carbon by the PKS. Genetic studies have so far failed to solve the "starter unit problem", but have thrown up some observations relevant to it. For example cloning of the *otc* minimal PKS (in the presence of the *act* KR) in *S. coelicolor* CH999 yielded a decaketide derived entirely from acetate (RM20: **20**), just as the *tcm* minimal PKS had done. This showed that the *otc* PKS can use an acetate starter unit, but left open the natural determination of the correct starter.[107] Similarly, nonaketides derived from nine acetate residues were produced by the minimal *fren* PKS in *S. coelicolor*, as described above. The daunorubicin PKS gene cluster in *S.* sp. C5 (Figure 5) includes a gene (*dpsC*) whose product resembles the *E. coli* ketosynthase III, specific for the first condensation in fatty acid biosynthesis, and a second gene (*dpsD*) whose product shows sequence similarities with acyl transferases, suggesting that they might be involved in the choice of the appropriate starter unit, but neither gene was required for the PKS to produce polyketide chains starting with propionate in the *S. coelicolor* expression system.[108] Elucidation of starter unit choice for the initiation of aromatic polyketide carbon chain assembly, and whether chain initiation might share steps with that of fatty acid synthesis in the same organism (see section VI.C), will probably require a more biochemical approach, aided by genetics, for its solution.

6. Predictions about Biosynthesis and Structure

An unexpected benefit of cloning recombinant sets of PKS subunits in *S. coelicolor* has been its use to help to solve biosynthetic questions in situations in which chemical approaches had proven difficult.

One example concerns the spore pigments of *Streptomyces* species, which have resisted attempts to isolate them for chemical characterization. Cloning and expression of the *whiE*, *sch*, and *cur* minimal PKS gene sets from *S. coelicolor*, *S. halstedii*, and *S.curacoi* respectively, alone or in combination with *act* PKS subunit genes, in *S. coelicolor* strain CH999,[97,105] has clearly established their products as C_{24} and (at least in this artificial situation) C_{22} polyketides, even though the complete structures of the spore pigments themselves are unknown.[109] Similarly, the finding of decaketide products for the cloned *mon* minimal PKS suggests that the unknown aromatic metabolite of *S. cinnamonensis* which it presumably makes is itself a decaketide.

A further example concerns the aureolic acid group of polyketides, such as mithramycin (**12**, Figure 3) produced by *Streptomyces argillaceus*, for which two possible routes for carbon chain assembly had been proposed: either a fusion between separate octa- and diketide chains, or direct assembly of a decaketide. Cloning of the *mtm* minimal PKS in *S. coelicolor* in the presence of the *act* KR produced the already identified decaketides RM20b (**23**) and RM20c (**24**), clearly establishing that the *mtm* PKS can synthesize C_{20} chains and suggesting this as the most plausible biosynthetic origin of the mithramycin carbon backbone.[110]

7. Design Rules for Novel Aromatic Polyketides

As described above, analysis of the polyketides produced by many recombinants carrying different combinations of PKS subunits led to a series of conclusions about PKS programming. This retrospective analysis could be converted into a prospective set of "design rules" with enough predictive power to generate novel molecules "to order", and at least three have already been made: SEK43 (**34**) and SEK26 (**35**),[98] as well as PK8 already mentioned.[96] These rules are summarized in Table 2.

G. *In Vitro* Studies

Although much has clearly been learned about the responsibility of different PKS subunits for various aspects of programming, the mechanisms involved are still largely unknown. Now that genetic studies have set the scene, biochemical approaches, often making use of genetically engineered strains, will make an increasingly important contribution. Significant steps toward the development of fully controlled *in vitro* systems have been taken by the expression of components of minimal PKSs in *E. coli*: the *tcm* KS, CLF, and ACP,[112,113] and the *act*, *gra*, *otc*, and *fren* ACPs.[114] A solution structure of the recombinant *act* ACP by NMR spectroscopy represents the first such structure for any PKS component.[115,116] A ground-breaking step toward the goal of reconstituting *in vitro* polyketide synthesis was taken by Shen and Hutchinson,[117] who obtained synthesis of tetracenomycin F2, a tetracenomycin C precursor, from added acetyl and malonyl CoA in a cell-free extract from an *S. glaucescens* recombinant strain in which the *tcm* KS, CLF, ACP, and aromatase/cyclase (TcmN) PKS subunits were overexpressed. When the ACP in the extract was biochemically depleted, synthesis was much reduced, as expected, but could be restored by adding back recombinant *tcm* ACP made in *E. coli*. More recently, recombinant TcmN protein was added back to an extract prepared from a strain lacking it, causing it to produce tetracenomycin F2 instead of (largely) SEK15.[118] Using cell-free preparations from *S. coelicolor* recombinants, Carreras *et al.* obtained synthesis of SEK4 and SEK4b from acetyl and malonyl CoA when the minimal PKS was present, and DMAC when the set of genes included not only those for these three PKS subunits but also for the KR, ARO, and CYC.[119] All these results are in striking agree-

Table 2. Design Rules for the Biosynthesis of Novel Aromatic Polyketides

structural feature	design rules and comments
carbon chain length	controlled by the minimal PKS (KS+CLF+ACP)
	ACPs are interchangeable (at least in the range C16-C24)
	the CLF is crucial to correct chain length, but heterologous KS+CLF combinations are often nonfunctional; therefore, choose homologous KS+CLF pairs for engineering novel molecules
	in the (unusual) cases of relaxed chain length control (*fren*, *whiE*), other PKS subunits can influence the choice (by one chain-extender unit): for example the *fren* minimal PKS alone made only C_{16} chains, with *tcm*N it made only C_{18} chains, and with the *act* KR it made a mixture
	CLFs presumably have a common ancestry with KSs and have diverged more from each other and from the common progenitor
	homologous KS and CLF *may* form a heterodimer
ketoreduction	a specific ketoreductase is required
	the *act* KR (the only KR to have been studied in detail) works on chains of at least C_{16}-C_{24}
	the *act* KR normally reduces at C-9 (occasionally at C-7); probably most other KRs have similar regiospecificity because their cognate natural products show C-9 ketoreduction
	of the sequenced PKS gene clusters, *tcm*, *whiE*, *sch*, and *cur* lack a KR
first ring cyclization	can be controlled by the minimal PKS, but regiospecificity can be influenced by other PKS subunits; for example the *tcm* minimal PKS alone made a mixture of C-7/C-12 and C-9/C-14 cyclized compounds, but in presence of *act* KR, ARO and CYC, C-9/C-14 cyclized compounds arose, and in *S. glaucescens*, only C-9/C-14 cyclized compounds; *tcmN* caused the *act* minimal PKS to form C-9/C-14 cyclized products instead of C-7/C-12
	in the presence of the *act* KR (and presumably others), regiospecificity of first ring cyclization depends on the position of ketoreduction: C-7/C-12 for C-9 reduction, C-5/C-10 for C-7 reduction
first ring aromatization	for unreduced molecules, this is uncatalyzed
	for reduced molecules, needs an aromatase; AROs like the *act*VII homologues are internally duplicated ("didomain" proteins), perhaps reflecting the need to extract two molecules of water; they are specific for chain length:[98] *gris* ARO works on C_{20}, C_{18}, C_{16}; *fren* ARO works on C_{18}, C_{16}; *act* ARO works on C_{16}
second ring cyclization	needs an appropriate cyclase (such as the *act*IV protein); such CYCs show some chain length specificity: e.g., *act* CYC works on C_{16} and C_{18} but not C_{20} chains; presumably CYCs for such longer chains remain to be discovered
	for sufficiently long unreduced chains with C-9/C-14 first ring, the minimal PKS (e.g., *tcm*) catalyzes second ring cyclization (C-7/C-16)
second ring aromatization	a "monodomain" ARO, the *tcmN* protein, is responsible for aromatization of the second ring of unreduced chains, for which the first ring aromatizes spontaneously
further cyclizations	the experiments performed so far with recombinant PKSs have not included proteins needed for *natural* cyclization reactions beyond the second ring, and often only the first; in these experiments, the free chain ends, when long enough, cyclize spontaneously by "chemical" rules of the kind previously indicated by biomimetic studies:[111] methyl ends give hemiketals and benzene rings; carboxyl ends give pyrones and may be decarboxylated if a free β-carboxyl exists; alternatively, aldol condensations may occur *between* ends, giving further carbocyclic rings
choice of starter unit	there is little information on this; the *dps* minimal PKS could make the correct choice of propionate starter in *S. coelicolor*, but the *otc* minimal PKS used acetate instead of the presumed natural starter

ment with the *in vivo* findings described above and augur very well for the further development of fully *in vitro* PKS systems and their use to elucidate PKS mechanisms.

III. Modular Polyketide Synthases for Macrolide Biosynthesis

As mentioned in section I, the programming mechanism for the complex or reduced polyketides, represented by the aglycons of the various classes of macrolides, is quite different from that for the aromatic polyketides. The first revelations about the modular structure of a macrolide PKS came from studies of erythromycin biosynthesis in *Saccharopolyspora* (formerly *Streptomyces*) *erythraea*, and this remains the type system for this class of PKSs. In this section, I briefly review the isolation of the genes for the erythromycin PKS and then touch on the genetics of some other macrolide PKSs (Table 3), before examining the evidence for and implications of the programming model itself.

A. Cloning and Sequencing of the Erythromycin PKS Genes

In 1982, Thompson *et al.*,[120] as part of a study aimed at the isolation of a series of *Streptomyces* self-resistance genes, described the cloning from *Sac. erythraea* of a DNA fragment that conferred resistance to erythromycin (**36**: Figure 12) on *S. lividans*. A single clone, pIJ23, was obtained, but it yielded so little DNA that the resistance gene (later named

Table 3. Cloning of Genes for Modular PKSs from Bacteria

host	polyketide (aglycon)	PKS genes	cloning strategy	evidence for cloning of correct genes[a]	module and gene organization[b] (nucleotide sequence accession numbers)	ref(s)
Sac. erythraea	erythromycin (36) (6-deoxyerythronolide B) (37)	*ery*	resistance, followed by walking and complementation	1,2,3,4,5	[S+2]+[2]+[2+TE] genetic order colinear with functional order (X56107, M63676–7)	120,123–132
S. fradiae	tylosin (38) (tylactone)	*tyl*	reverse genetics for a tailoring step, followed by walking and complementation	1,2	[?] genetic order colinear with functional order? (Not deposited)	136,137,140
S. ambofaciens	spiramycin (39) (platenolide)	*srm*	resistance, followed by walking and complementation	1,5	[S+2]+[1]+[2]+[1]+[1+TE] genetic order colinear with functional order? (Not deposited?)	140,141
S. thermotolerans	carbomycin (platenolide)	*car*	resistance, followed by walking and complementation	1	–	142
S. antibioticus	Oleandomycin (40) (oleandolide)	*ole*	*ery* and resistance gene probes	–	[?]+[2+TE] only one protein with two modules sequenced (L09654)	143
S. avermitilis	avermectins (41)	*avr*	complementation of blocked mutants, followed by walking	1,2	2×[3]+? genetic order not colinear with functional order: two converging sets of modules	144–146
S. cyanogriseus	nemadectin	*nem*	*avr* probes	2	[?] genetic order not colinear with functional order: two converging sets of modules	146
S. sp. FR-008	FR-008 (candicidin aglycon) (42)		paba-synthase followed by *ery* probes	2	–	150
S. hygroscopicus	rapamycin (43)	*rap*	*ery* probes	2,5	[6]+[4]+[4] genetic order not colinear with functional order (X86780)	147
S. sp. MA 6548	FK506 (44)	*fkb*	reverse genetics for a tailoring enzyme, followed by walking and sequencing	2	[4]+? only one gene with four modules sequenced	148–149
Sorangium cellulosum	Soraphen A (45)		*gra* (Type II PKS) probe	2	only one module and part of a second sequenced (U24241)	151

[a] Key: 1, complementation of pathway-blocked mutants; 2, gene disruption; 3, production of polyketide after transfer of cloned genes to *S. coelicolor*; 4, production of relevant compounds in recombinant strains; 5, agreement of module number, starting module, and terminating domain, and level of reduction with those predicted by chemical structure. [b] Square brackets show the modular content of protein subunits (e.g., [S+2] represents a subunit carrying a starter module and two chain extender modules), and arrows show the direction of transcription of the genes.

ermE) had to be rescued by subcloning to yield pIJ43. This plasmid had a difficult birth—apart from being very nearly lost, it suffered some rearrangements of vector sequences, and lacked the native *ermE* promoter[121]—but it turned out to have a bright future. The plasmid was dispatched to B. Weisblum for sequencing of the *ermE* gene,[122] and it, or a derivative of it, was sent to the laboratories of C. R. Hutchinson, P. F. Leadlay, L. Katz, and R. H. Baltz, where the *ermE* gene was used as a hybridization probe to clone genes for erythromycin biosynthesis from the genome of *Sac. erythraea*. DNA fragments isolated by genomic walking from *ermE* were sequenced, and used in gene disruption and complementation experiments, by these laboratories. In this way, numerous genes encoding tailoring steps in erythromycin biosynthesis were found. One lay on the 5′ side of *ermE* (to the left as the gene cluster is conventionally drawn), and many on the 3′ (right) side.[123–127] A key step was the identification of a segment of DNA that complemented the *eryA* class of mutations, which were blocked in biosynthesis of the aglycon (polyketide) moiety of erythromycin (6-deoxyerythronolide B: **37**), some 12 kb downstream of *ermE*;[128] even more significantly, as it turned out, another segment of DNA that hybridized to the one that complemented the *eryA* mutations was found some 35 kb downstream of *ermE*, implying that *eryA* was a complex locus covering a considerable stretch of DNA.[129] The stage was thus set for the sequencing of *eryA* in the Leadlay and Katz laboratories. The results were spectacular: the sequence revealed three unusually large genes, each encoding a protein carrying two modules of PKS active sites, with each module resembling in its sequence and organization a vertebrate fatty acid synthase;[130–132] these are the

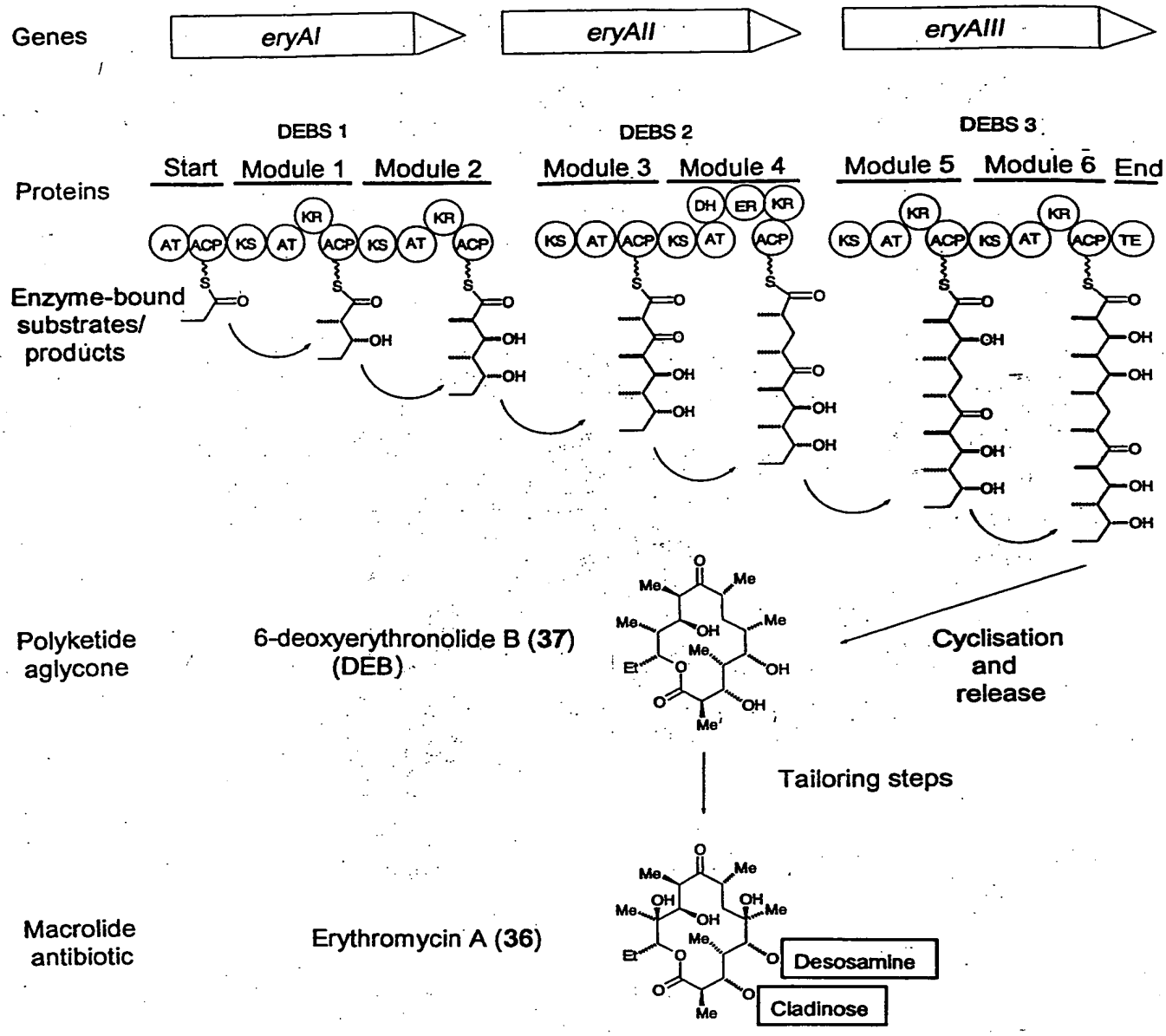


Figure 12. The modular erythromycin PKS and the "assembly-line" model of biosynthesis leading to 6-deoxyerythronolide B, and on to erythromycin A.

three proteins that are now known as 6-deoxyerythronolide B synthase (DEBS) 1, 2, and 3 (Figure 12). The three genes were seen to be arranged in the same order as the proposed sequence of action of the six modules of active sites that they encoded: DEBS1, carrying a short starter module and chain extender modules 1 and 2; followed by DEBS2, carrying modules 3 and 4; and then by DEBS3, carrying modules 5 and 6 together with a final thioesterase domain for carbon chain release. The existence of these three hypothetical proteins was confirmed in a biochemical *tour-de-force* in which they were revealed by Western blotting and N-terminal sequencing.[133]

B. Genes for Other Macrolide PKSs

1. Tylosin

Tylosin (**38**: Figure 13), produced by *Streptomyces fradiae*, was one of the first antibiotics for which a comprehensive set of blocked mutants was isolated. These were used to help to define the biosynthetic pathway[134] and to adumbrate the potential for genetic engineering in manipulating it, either to increase the productivity of the fermentation or to generate novel antibiotics.[135] Whereas erythromycin is a 14-membered macrolide, tylosin has a 16-membered ring structure. The gene (*tylF*) for the final biosynthetic step—O-methylation of the tylosin precursor, macrocin—was cloned by reverse genetics from the protein sequence of the enzyme, and specific segments of surrounding DNA were found to complement eight other classes of blocked mutants, but not the *tylG* mutants that were presumed to be defective in the tylosin PKS.[136] Soon the genetic and physical map of the cluster was extended to cover a ~90 kb segment bounded by two resistance genes.[137] It included DNA segments that complemented nearly all the known classes of blocked mutants, but only one member of a large set of *tylG* mutants. This

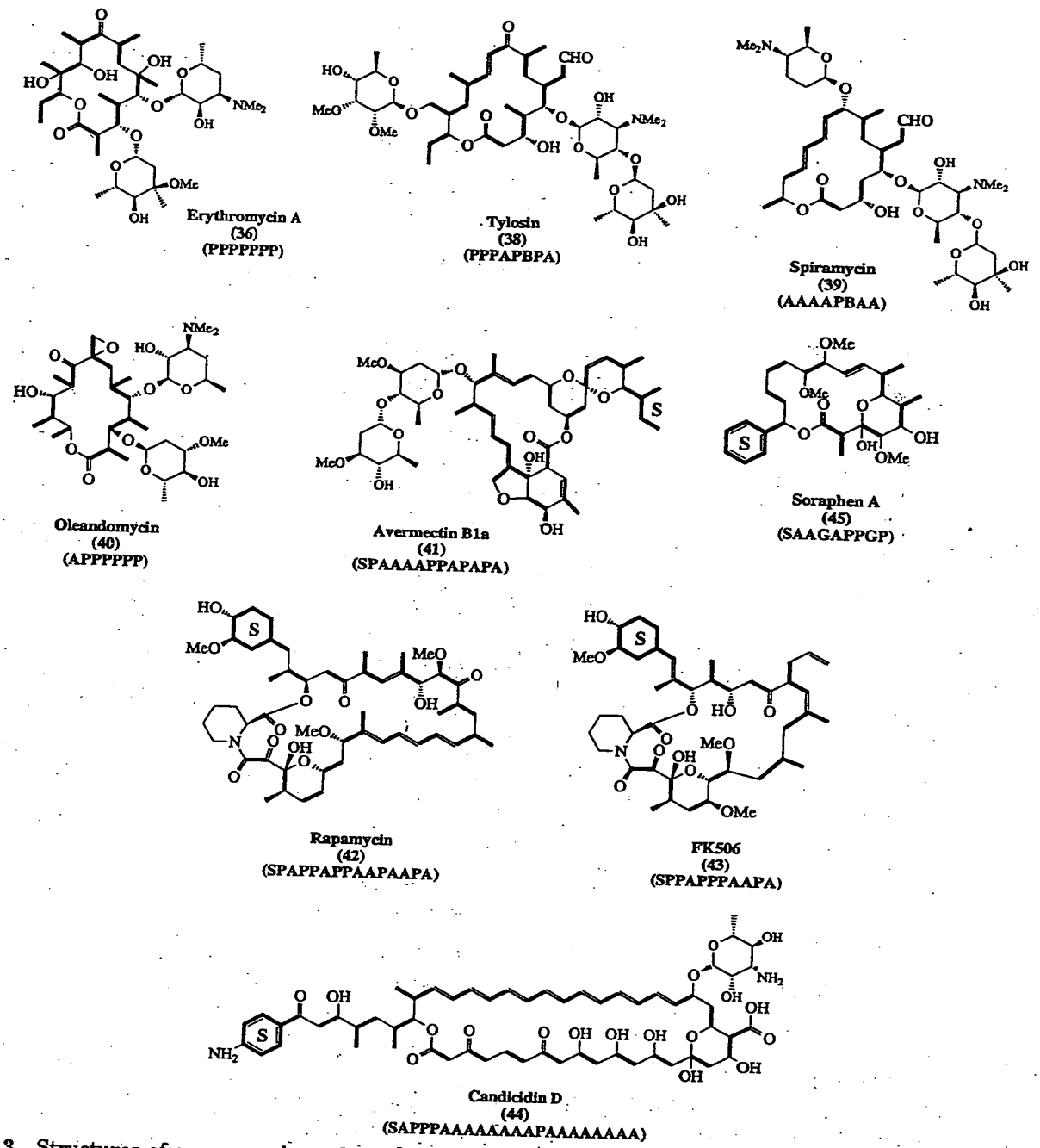


Figure 13. Structures of some complex polyketides produced by modular PKSs in bacteria. In each structure, heavy bonds indicate the carbons derived from acetate (A), propionate (P), butyrate (B), glycerol (G), or from branched chain or cyclic starter units (S) that go to build the polyketide chains; the sequence of letters below the name of each compound shows the sequence of residues that would build each polyketide chain. (Figure kindly drawn by C. Surti.)

latter segment was adjacent to a ~30 kb stretch of DNA that included a series of (imperfect) direct repeats and a segment, named an Amplifiable Unit of DNA (AUD), that was subject to repeated amplification and deletion events caused presumably by recombination involving the repeats.[138,139] Some of these deletions and amplifications had earlier been associated with a *tylG* phenotype. This mysterious situation—which not surprisingly puzzled and frustrated the Lilly group for several years—was clarified as soon as the modular structure of the erythromycin PKS was announced: the 30 kb stretch of the tylosin biosynthetic gene cluster represented a complex *tylG* locus encoding a modular tylosin PKS, and recombination between the conserved sequences in the various modules was the likely explanation for the observed amplification/deletion events. Why this has not been observed for the corresponding *eryA* locus (or those for other cloned modular PKSs) is unknown. Perhaps the *tylG* modules retain more DNA sequence similarity than the corresponding *eryA* modules; or perhaps the extent of nucleotide sequence identity required for homologous recombination in *S. fradiae* is less than in *Sac. erythraea*, so that the recombination between different modules that would trigger amplification occurs at a noticeable frequency in the former host but not in the latter. Unfortunately, although the sequences of the tylosin PKS genes have

apparently been determined by the Lilly group (referred to in ref 140), they are not yet in the public domain.

2. Spiramycin and Carbomycin

Spiramycin (39) and carbomycin, produced by *Streptomyces ambofaciens* and *Streptomyces thermotolerans* respectively, are two further 16-membered macrolide antibiotics both derived from the same primary polyketide metabolite, platenolide. For both, cloning of one or more resistance genes in a heterologous *Streptomyces* host led to identification of linked biosynthetic genes that were revealed by complementation of blocked mutants with segments of the cloned DNA.[141,142] Complete PKS gene sequences, at least for the spiramycin PKS (*srmG*), have been obtained by the Lilly group. The sequences themselves do not appear to be in the public domain but a summary of the deduced modular structure of the *srmG*-encoded PKS has been published.[140] The seven modules needed to assemble the octaketide are distributed over five PKS subunits: two bimodular proteins (like the three DEBS proteins), the first including a starter module, and three unimodular proteins, the last carrying a thioesterase release domain (Table 3).

3. Oleandomycin

One open reading frame that would encode a protein carrying two modules of a PKS was cloned from the producer of another 14-membered macrolide aglycon, that of oleandomycin (40) produced by *Streptomyces antibioticus*, by the use of a DEBS3 PKS probe and an oleandomycin resistance gene probe.[143] No definite proof of its involvement in oleandomycin biosynthesis was obtained, but the close linkage of a specific resistance gene is suggestive, and the presence of a thioesterase domain at the C-terminus of the presumptive product of the gene is consistent with its encoding the final two modules of the presumably six-module oleandomycin PKS.

4. Avermectin and Nemadectin

A biosynthetic gene cluster for the avermectins (41), important antiparasitic polyketide macrocyclic lactones produced by *Streptomyces avermitilis*, was isolated by complementing mutants blocked in tailoring steps of the biosynthetic pathway: for O-methylation and glycosylation of the polyketide moiety.[144] Chromosome walking between the two complementing regions, and gene disruptions, identified a 165 kb region containing the *avr* gene cluster. Centrally located in the cluster, DNA encoding a modular PKS was found by hybridization and limited DNA sequencing.[145] From this it was deduced that the genes for 12 PKS modules needed to assemble the avermectin aglycon are organized differently from those encoding the DEBS modules in two respects: (1) the genes are not colinear with the functional order of action of the proteins, but instead the DNA encoding the 12 modules forms two converging sequences, each apparently encoding six modules; and (2) two of the genes appeared to encode proteins each carrying three modules, instead of each encoding two modules as in DEBS (distribution of the remaining modules is unknown).[146] Use of *avr* PKS sequences as probes against DNA of *Streptomyces cyanogriseus*, the producer of nemadectin, with the same aglycon structure as avermectin, led to the isolation of genes (*nem*) encoding another modular PKS, and this was proved by gene disruption to be involved in nemadectin biosynthesis.[146] Hopefully, sequence information on the *avr* and *nem* PKS genes will become available in due course.

5. Rapamycin and FK506

Rapamycin (42) is an important immunosuppressant made by *Streptomyces hygroscopicus*. The PKS genes, and most or all of the genes encoding tailoring steps in rapamycin biosynthesis, were cloned by the use of DEBS PKS probes, followed by chromosome walking and sequencing of 107 kb of DNA.[147] The PKS consists of three RAPS proteins carrying the active sites needed for 14 condensation cycles; one protein consists of six modules and the two others contain four modules each. In contrast to the erythromycin case, the genetic order is not the same as the functional order of activity of the three proteins (Table 3), and typical starter and thioesterase domains are not seen, consistent with the different structures of the rapamycin and erythromycin polyketides: rapamycin starts with incorporation of a cyclohexane carboxylic acid unit and ends with incorporation of a pipecolic acid unit. Gene disruption has been used to prove the involvement of these genes (but not of two other DNA regions from the same *S. hygroscopicus* strain, each presumptively encoding a modular PKS) in rapamycin biosynthesis.[147a]

FK506 (43) is another important immunosuppressant, made by *Streptomyces* sp. MA 6548. Reverse genetics led to the cloning of a gene for a tailoring enzyme, an O-methyltransferase, which was shown by gene disruption to be essential for FK506 biosynthesis.[148] Nearby, sequencing revealed a gene for a cytochrome P450 hydroxylase, responsible for another biosynthetic step,[148] and a large open reading frame which would encode a modular PKS subunit carrying four modules of active sites, and closely resembling one of the RAPS proteins.[149]

6. Candicidin

Candicidin (44) is a member of the polyene macrolide class of complex antifungal polyketides. A related compound, FR-008, with an identical aglycon to that of candicidin, is made by *Streptomyces* sp. FR-008. A gene cluster involved in FR-008 biosynthesis was isolated by hybridization, initially using as probe a gene involved in biosynthesis of the *p*-aminobenzoic acid-derived starter unit; later the DEBS2 gene was used as a PKS probe, and finally probes specific for ACP, KS and AT motifs from the DEBS PKS.[150] The hybridization patterns with these last probes revealed that DNA encoding a modular PKS extended over 105 kb. This would be an appropriate length to encode 21 PKS modules of average length ~5 kb, and this is the number of condensations required for synthesis of the FR-008 aglycon. The finding is significant in implying a one-to-one relationship

between modules and rounds of condensation, even in a polyene with a high degree of chemical repetition represented by the seven successive double bonds all associated with the insertion of acetate residues. *A priori*, this part of the polyketide might have been assembled by the iterative activity of a single module, but this result makes this idea unlikely.

7. Soraphen

Soraphen A (**45**), produced by a myxobacterium, *Sorangium cellulosum*, is made by the first example of a functional modular PKS so far known outside of the actinomycetes. Interestingly, it was possible to clone DNA encoding part of the PKS by the use of a probe encoding part of one of the type II, nonmodular PKSs, for granaticin biosynthesis (see section II).[151] Gene disruption proved the involvement of the cloned DNA in soraphen biosynthesis, and sequencing has so far revealed part of a gene encoding one complete module of PKS active sites and an incomplete second module.

C. The Programming Model and Its Proof by Mutant and Recombinant Construction

Initial evidence for the assembly-line model for programming of the erythromycin PKS (Figure 12) was provided by the sequence itself.[131] Not only did the six modules of putative catalytic sites correspond in number to the six condensations needed to build the erythromycin heptaketide, but special features of specific modules could be related to their proposed functions: DEBS1 had extra N-terminal AT and ACP domains, before module 1, which would function in the loading of the propionyl CoA starter unit (only later was this part of the protein designated as a separate "starter" or "loading" module), and DEBS3 was unique in carrying a putative thioesterase domain after module 6 for hydrolysis of the final thioester bond between the completed polyketide chain and the 4′-phosphopantetheine prosthetic group of the last ACP domain to release the carbon chain; module 3 lacked all three reductive functions (KR, DH, and ER), agreeing with the presence of an unreduced keto group after the third condensation, while module 4 was unique in carrying candidate domains for all three such functions, as expected in view of the reduction of the keto group right through to a methylene after the fourth condensation. Experimental evidence was soon provided by two successful domain inactivation experiments: when the active site of the putative KR in module 5 was deleted, a polyketide with an unreduced keto group after the fifth condensation (**46**: Figure 14) was isolated;[131] and when the putative ER in module 4 was mutated, a double bond appeared in the final product (**47**), as expected from a failure of enoyl reduction after the fourth condensation.[152]

These results were indeed impressive in establishing the assembly line model. Further strong support for the model was provided by the construction of recombinants carrying reduced numbers of DEBS modules, which were found to produce various "minilactones" of appropriate structure. One set of experiments (here referred to as the "Stanford experiments", but involving a collaboration of the Khosla with the Katz and Cane laboratories) stemmed from the remarkable feat of engineering expression of the entire set of DEBS proteins in the host−vector system (*S. coelicolor* CH999/pRM5) that had already proved so successful for analyzing type II PKS structure−function relations (section II): the resulting recombinant produced significant quantities of 6-deoxyerythronolide B, together with an analogue with an acetate starter (**48**: Figure 14).[153] When the DEBS1 protein was expressed alone, a triketide lactone (**49**) arose,[154] which was precisely the expected product of modules 1 and 2 (and had earlier been reported as a minor product in the mutant deleted for the KR of module 5:[155] it increased in quantity when the thioesterase domain, which normally forms the C-terminus of DEBS3, was added to DEBS1.[156] This construct (DEBS1+TE) was the one that had already been engineered in exchange for the normal DEBS1 in a *Sac. erythraea* host deleted for DEBS2 and most of DEBS3 in the "Cambridge experiments", which also yielded the triketide lactone (**49**).[157] (In both the Stanford and the Cambridge experiments, recombinants carrying DEBS1+TE in the *S. coelicolor* surrogate host also produced an unnatural triketide lactone (**50**) with an acetate instead of a propionate starter.[156,158]) The production of tetraketide (**51** and **52**) and hexaketide (**53**) lactones by *S. coelicolor* recombinants carrying modules 1−3 or 1−5 (together with the TE) provided further evidence for the programming model,[156,159] and for the remarkable degree of functional independence of which the various modules are capable.

Most recently, success has been achieved in attempts to generate unnatural natural products by domain *replacement* rather than deletion or mutagenesis. In one of these, advantage was taken of the key finding, from the sequencing of the rapamycin PKS genes, of a difference in the consensus sequence for the seven AT domains that would introduce acetate extender units (from malonyl CoA) from that of the seven domains that would introduce propionate extenders (from methylmalonyl CoA)[157] (the latter consensus was also shared by the AT domains in DEBS, which handle methylmalonate). Again exploiting the CH999/pRM5 expression system, the methylmalonyl transferase domain of DEBS1 module 1 was replaced by the malonyl transferase domain of RAPS1 module 2, to generate a functional PKS that produced novel triketide lactones (**54** and **55**) of the predicted structure, with incorporation of an acetate instead of a propionate residue as the first chain-extender unit.[160] In a second example of domain swapping, the starter module of the spiramycin aglycon PKS, which makes platenolide in *S. ambofaciens* by incorporating an acetyl unit, was replaced by the starter module for the *S. fradiae* tylactone PKS, which normally incorporates a propionyl starter: the result was the predicted methyl platenolide.[140] Another example of starter module swapping involved exchange of the starter module of DEBS1+TE (cloned in *S. coelicolor*) by the starter domain of the avermectin PKS; the recombinant made the predicted triketide lactones that start with the branched-chain residues characteristic of the natural avermectins (**41**: Figure 13).[161] Finally, a

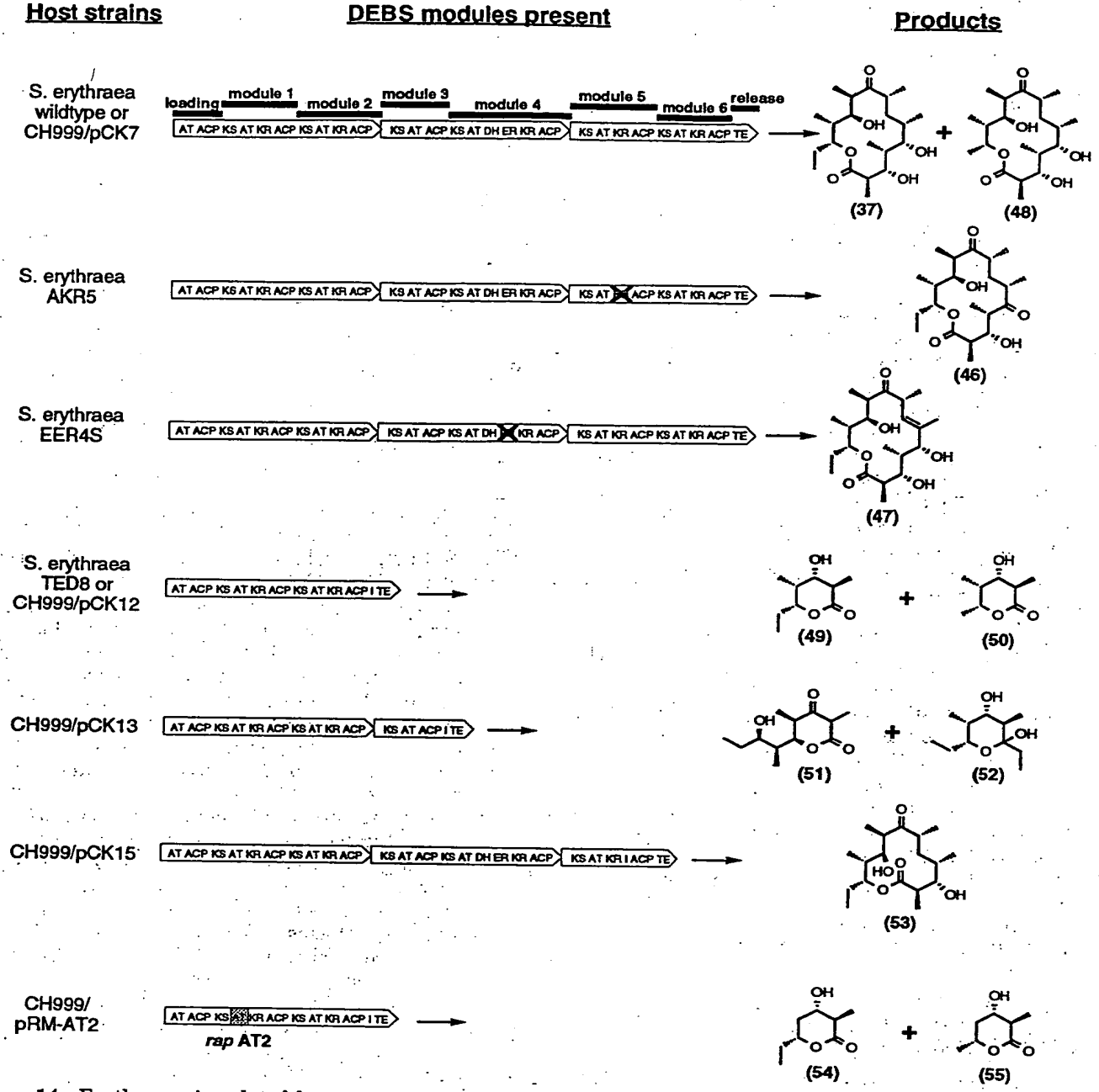


Figure 14. Erythromycin-related lactones made by the complete (6-module) 6-deoxyerythronolide B synthase (DEBS) (Figure 12), or artificially constructed synthases with mutations, reduced numbers of modules, or a domain substitution, in either *Sac. erythraea* or *S. coelicolor* CH999. Note that the lactones with an acetate starter (**48**, **50**, or **55**) are found, together with metabolites with the usual propionate starter (**37**, **49** or **54**, respectively), in the *S. coelicolor* host but not in *Sac. erythraea*. References: **37** and **48**;[153] **46**;[131] **47**;[152] **49** and **50**;[154,157,158] **51** and **52**;[159] **53**;[156] **54** and **55**.[160] (Figure kindly drawn by C. Khosla.)

very important development was the *addition* of a dehydatase domain (from the rapamycin PKS) to DEBS1 module 2, to yield a product with a "gain of function" double bond in the expected position.[161a]

D. *In Vitro* Studies and Model Building

Genetic approaches have been crucial in deducing the basic features of the programming rules for the modular type I PKSs, but understanding how these rules are interpreted by the cellular machinery will have an increasingly large biochemical input. Some very significant steps have already been taken. For example, in a beautiful *in vitro* experiment with purified DEBS1, DEBS2, and DEBS3 proteins, it was found that only the 2*S*, and not the 2*R* stereoisomer of ^{14}C-labeled methylmalonyl CoA was attached to all the AT sites of the PKS subunits, providing evidence for the use of only the 2*S* stereoisomer in chain extension (presumably with racemization after incorporation of the second, fifth, and sixth extension units to give the *R* configuration found in 6-deoxyerythronolide B), rather than selective use of (2*S*)- and (2*R*)-methylmalonyl CoA as chain extenders.[162]

Both the Stanford and the Cambridge groups have taken the important step of developing active *in vitro* systems, using recombinant DEBS proteins; these used either the complete, natural, three-component PKS[163] or the DEBS1+TE engineered protein.[163-166] Interestingly, the recombinant enzymes, *in vitro*, showed considerable permissiveness for starter units, extending beyond just propionyl CoA or acetyl CoA.[165]

Biochemical and genetic studies have also begun to define important features of the three-dimensional structure of the native DEBS protein. A very interesting model for this, proposed by the Cambridge group largely from the results of rigorous chemical cross-linking experiments, is a double-helical homodimer in which the domains found in all of the modules (KS, AT, and ACP) form the core of the helix and the reductive domains form loops of various lengths depending on the presence of KR, KR+DH, or KR+DH+ER domains in a specific module.[167] An attractive feature of the model is its overall head-to-head/tail-to-tail organization, which would nicely accommodate the additions and/or deletions of modules that are a logical consequence of the presumed evolutionary relatedness of the various modular PKSs (section VI). A model proposed by the Stanford group, from mutant complementation experiments, also has an *overall* head-to-head/tail-to-tail dimeric structure, but within each module the two identical polypeptide chains would lie head-to-tail.[168] This arrangement, in which there would be two equivalent clusters of active sites in each module, with each protein subunit contributing certain sites to each cluster (for example the KS to one and the ACP to the other), not only arises elegantly from the complementation results, but is reminiscent of the head-to-tail homodimeric structure of the ("unimodular") vertebrate fatty acid synthase,[169,170] to which the PKS modules are doubtless phylogenetically related (see section VI). Especially in view of the fact that the helical aspect of the Cambridge model is arbitrary,[167] the two models are in fact essentially equivalent topologically.

IV. 6-Methylsalicylic Acid Synthase and Other Fungal Polyketide Synthases

6-Methylsalicylic acid synthase (6-MSAS) is the classical PKS, which has been studied from a biochemical standpoint ever since synthesis of 6-MSA (**56**: Figure 15) from acetyl CoA and malonyl CoA by an enzyme extract from *Penicillium patulum* was reported in pioneering work from the Lynen laboratory.[171] The gene was cloned by screening an expression library in *E. coli* with an antibody prepared against the purified protein.[172] Sequencing revealed a single open reading frame (interrupted by a short intron); most significantly, the four active sites identified in the sequence of the encoded protein (KS, AT, KR, and ACP) resembled the corresponding sites in rat FAS, and were colinear in the two synthases.[172] This provided early evidence for a presumptive phylogenetic link between eukaryotic PKS and FAS genes, and detailed biochemical similarities tended to support this idea.[173] Apart from furnishing crucial information on the primary structure of the enzyme, genetic contributions to understanding 6-MSAS function are so far few. They should be aided by the expression of the engineered gene in the *S. coelicolor* CH999 host, in which it gave rise to significant quantities of 6-MSA.[174]

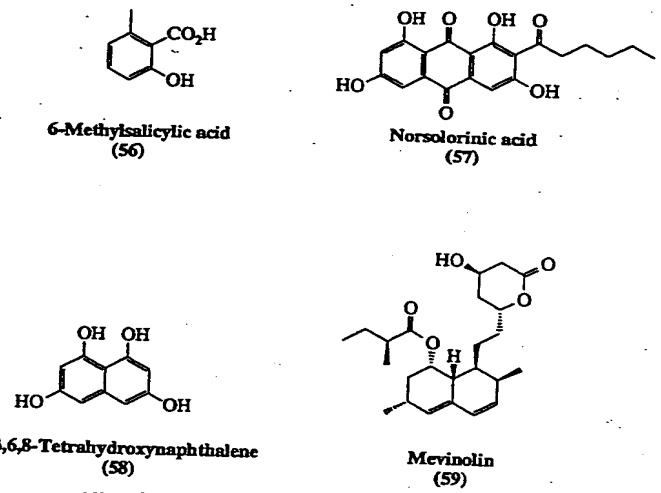


Figure 15. Structures of some aromatic polyketides produced by type I (iterative) PKSs in fungi. (Figure kindly drawn by C. Surti.)

Genes for several other type I PKSs from filamentous fungi have subsequently been cloned and sequenced. These include PKSs for an asexual spore pigment of unknown structure in *Aspergillus nidulans*;[175] the decaketide norsolorinic acid (**57**), an intermediate of aflatoxin biosynthesis in *Aspergillus parasiticus*;[176-178] the pentaketide precursor (1,3,6,8-tetrahydroxynaphthalene: **58**) of fungal melanins in *Alternaria alternata*[179] and *Colletotrichum lagenarium*;[180] and the nonaketide mevinolin (lovastatin) (**59**) in *Aspergillus terreus*.[181] The deduced products of these PKS genes, like 6-MSAS, all show an overall structural resemblance to the vertebrate FAS, with significant absences of particular domains. Thus most of the synthases lack a C-terminal thioesterase domain (presumably release of the carbon chain from the synthase occurs by a mechanism different from a simple catalyzed hydrolysis of the final thioester bond); and all except the mevinolin PKS lack the domains for the three reductive functions, which are superfluous to the biosynthesis of unreduced polyketides.

Other features of interest have been noted in particular cases: for example a methyltransferase domain (needed for C-6 methylation) forms a covalent part of the structure of the mevinolin PKS. A feature of great genetic interest is the finding of close linkage of the genes for the various steps in the biosynthesis of these fungal metabolites; the most striking example is a cluster of 25 coregulated genes for biosynthesis of the aflatoxin-related metabolite sterigmatocystin (elaborated from norsolorinic acid) in *A. nidulans*.[182] Such clustering, which is characteristic of both primary and secondary metabolites in bacteria, is not normally found for primary metabolic pathways in fungi.

In an interesting recent turn of events, the aflatoxin (or related) gene clusters in *A. parasiticus* and *A. nidulans* have been found to encode two polypeptide chains typical of a fungal FAS, in which the

catalytic domains occur in a different order from those of the vertebrate FAS (Figure 2).[183-185] This presumed FAS is proposed to be dedicated to synthesis of a hexanoyl residue which, rather than acetate, would prime norsolorinic acid biosynthesis, and is distinct from the FAS used to make the lipids of the cell.[185] Perhaps this can be claimed as another example of genetic analysis helping to solve a problem in bioorganic chemistry—revealing that norsolorinic acid would be an octaketide rather than a decaketide[186]—to put with those arising from the manipulation of type II aromatic PKS genes (section II.F.6).

V. The Chalcone and Stilbene Synthase Superfamily from Higher Plants

Chalcone synthase (CHS) is a PKS by definition because it catalyzes the linking of acyl CoA subunits by repetitive decarboxylative condensations that is the hallmark of polyketide and fatty acid synthesis. However, there are significant differences in the biochemistry of the process, notably the participation of free CoA esters as substrates without the involvement of 4'-phosphopantetheine arms carried on acyl carrier proteins; and sequence comparisons suggest that CHS is phylogenetically distinct from all other groups of PKSs and all known FASs, with no significant overall resemblance of amino acid sequence, and a different motif surrounding the active-site cysteine.[187] CHS is a remarkable enzyme because, as a homodimer of a modest-sized polypeptide chain (only 43 kd), it selects coumaroyl CoA as starter, carries out three successive extensions using malonyl CoA (with no reduction of β-keto groups), and releases the resulting tetraketide to cyclize to naringenin chalcone (**60**: Figure 16), the precursor for the anthocyanin pigments and other flavonoids of plants.

The CHS gene from parsley (*Petroselinum hortense*) was the first to be cloned, as a cDNA by taking advantage of the abundance of its transcript in cultured cells.[188] It can therefore be regarded as the founding member of what has now grown to be a large superfamily of genes for these specialized PKSs found, so far exclusively, in higher plants. The family includes not only CHSs from many Angiosperm families and from Gymnosperms, but also related enzymes, the stilbene synthases (STSs), which share more than 65% amino acid sequence identity with CHSs, and build the same nascent tetraketide, but release it with decarboxylation and a different fold compared with chalcone to generate stilbene (**61**) (Figure 16).[189] Very interestingly from a genetic standpoint, a phylogenetic analysis of more than 30 CHS and several STS sequences showed a greater resemblance between CHS and STS sequences from related plants than among CHSs or STSs, suggesting that STSs have evolved from CHSs more than once; indeed it was possible to reproduce such changes in specificity by site-directed mutagenesis.[190] Apart from typical CHSs and STSs, the superfamily includes enzymes that generate other types of tetraketides that differ in the choice of starter unit, such as the bitter acids of hops (*Humulus lupulus*), which start with isovaleryl CoA or isobutyryl CoA.[191]

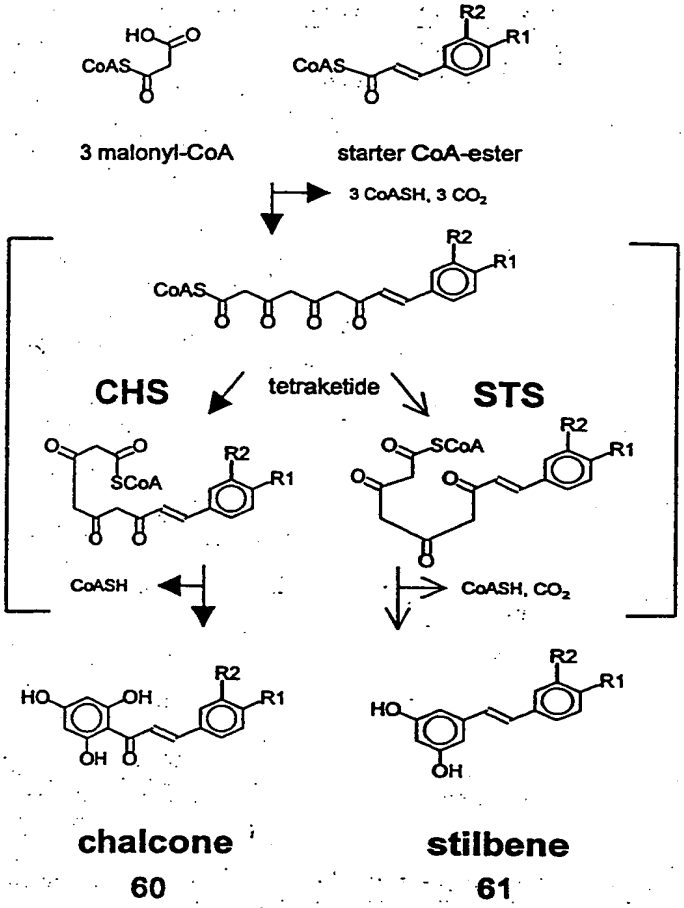


Figure 16. Reactions catalyzed by plant chalcone synthases (CHS) and stilbene synthases (STS) and structures of their "classical" products, naringenin chalcone (**60**) and stilbene (**61**). Note that both synthases use as starter a CoA ester from the phenylpropanoid pathway (such as coumaroyl CoA where R1 = —OH and R2 = —H)) and three malonyl CoA extenders to generate a linear tetraketide, which folds differently to connect different carbon atoms (solid or open arrows) to produce either **60** or **61** (the latter with decarboxylation). (Reproduced with permission from ref 187a. Copyright Elsevier Science Ltd., The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK. Figure kindly supplied by J. Schröder.)

Genetic manipulation has played a significant role in analysing CHS and STS function. Apart from site-directed mutagenesis to confirm the active-site cysteine and to interconvert the two classes of enzyme, just mentioned, elegant *in vitro* mutant complementation analysis has suggested that although the native enzymes are homodimers, each monomer can perform all three condensation reactons, but complementation between two inactive monomers to form an active enzyme suggests that the two monomers cooperate in other ways to generate product;[192] this might contrast with the situation for the vertebrate FAS (and perhaps the DEBS modules, section III.D), in which two active centers on the homodimer arise by obligate cooperation between complementary sites on the two subunits.[169,170] Much work has also been carried out on the transcriptional control of CHS (see the review by Martin[193] for an account of these studies and of the various biological roles of the plant metabolites produced by the CHS).

Table 4. Possible Genetic Events in the Evolution of FASs and PKSs

genetic events[a]	consequences; selective pressure; possible examples
mutation and recombination	origin of new specificities; to increase efficiency and versatility; e.g., origin of STS from CHS
duplication	origin of FASs from PKSs (or *vice versa*), and of diverse PKSs from each other; to increase chemical versatility
	origin of multiple modules, giving rise to "assembly line" programming; to increase chemical complexity of polyketides
	origin of CLF from primordial KS (Figure 9), perhaps generating heterodimeric condensing enzyme from homodimeric progenitor; to better control chain length?
	origin of didomain aromatase from monodomain progenitor (Figure 10); to handle reduced carbocyclic rings more efficiently?
	origin of extra FASs (or some subunits thereof); to provide specialized secondary metabolites; e.g., aflatoxins, *nod* factors
deletion	deletion of whole modules to shorten macrolide chains; to diversify product structure; e.g., hygrolidin vs conconamycin, or methymycin vs erythromycin[205]
	deletion of one, two, or three reductive domains (ER, DH, KR) from specific modules; to diversify functionality of macrolide chain
fusion	origin of type I from type II synthases (at least two fusion cycles needed to give rise to vertebrate and fungal synthases); to increase catalytic efficiency and/or to facilitate equimolar synthesis of catalytic sites in the absence of operons
	origin of multimodular proteins from unimodular progenitors, converting noncovalent to covalent joints in modular PKS; to increase fidelity of building the "assembly line"?
recruitment	of reductive cycle enzymes (KR, DH, ER), converting a primordial PKS into a FAS; to generate lipids for membrane biogenesis
	of cyclases and aromatases; to aid controlled ring formation in aromatic polyketides
	of loading functions; to diversify structure of starter units
	of release functions; to diversify modifications of nascent polyketide chains after assembly
	of group transfers occurring during chain building; to diversify chemical structure; e.g., methylation in mevinolin
horizontal transfer	acquisition of genes from a donor other than by inheritance from an ancestor; to accelerate adaptation; e.g., fungal PKS or FAS, or type I modules between bacteria (myxobacteria vs actinomycetes?)

[a] Sometimes the event could be the converse of the one illustrated: e.g., addition vs deletion of modules or domains.

VI. Phylogeny, Horizontal Gene Transfer, and Developmental Crosstalk: Some Speculations

A. Phylogenetic Relationships between Synthases

A bonus from the considerable recent interest in the primary structures of the proteins that make up the FASs and PKSs of a wide range of organisms is the creation of a large database of sequences that is potentially available for phylogenetic analysis. Already this has given rise to some interesting evolutionary speculations.

As discussed in section V, both sequence and biochemical data indicate that plant chalcone and stilbene synthases very likely represent a family of related enzymes with a separate origin from all the other enzymes that build carbon chains by catalyzing repeated decarboxylative condensations between small carboxylic acid residues. However, sequence comparisons suggest that all these other PKSs and FASs could well have had a common origin. It is possible to construct a relatedness tree that includes all the condensing enzyme subunits of the type II synthases, as well as the homologous domains of the type I enzymes;[194,195] the same can be done with the ACPs.[196] However, there are obstacles to carrying out a proper phylogenetic analysis of the data. One problem is the wide diversity of sequences; on the hypothesis of a common origin, this would reflect the vast evolutionary time scale over which sequence divergence has occurred. The result is that any tree that includes all the sequences is likely to be rather arbitrary in the relationships between its deeper branches. Another is the inevitably subjective choice of the boundaries between the functional domains of the type I enzymes. There may well be a functional need for "spacer" regions between the various catalytic sites, to allow them to assume the correct three-dimensional relationships with other sites in the same polypeptide chain or with partner chains in a dimeric or multimeric structure; in these regions the constraints are therefore likely to be geometric rather than involving specific amino acid residues, so the degree of conservation of sequence is likely to differ for domain and interdomain regions.[132] Depending on where the sequences are "cut" for inclusion in the analysis, a different tree will therefore result; and differences in detail between trees generated with, for example, KS or ACP sequences may at least in part reflect these factors. Thus a general question that arises in all phylogenetic sequence analysis—to what extent do amino acid sequence similarities (or differences) reflect functional constraints, rather than evolutionary relatedness (or divergence)—is particularly difficult to deal with.

In spite of such problems, which hopefully will not permanently stand in the way of a rigorous phylogenetic analysis, it is interesting to speculate a little on the genetic processes that would have occurred on the hypothesis of a common origin for the PKSs and FASs (Table 4) and to point out some of the anecdotal evidence for them. Meanwhile, it is relevant to note an observation, which could of course

be overturned at any moment by further research, that fatty acids and polyketides are apparently absent from the Archaea (their membranes contain isoprenoid ethers instead of the fatty acid-containing phospholipids that are found in the other two branches of the living world). The recent publication of the complete genome sequence of an Archaeon agrees with the failure to detect polyketide or fatty acid metabolites in members of the Archaea: no recognizable KS is seen in the sequence of the *Methanococcus jannaschii* genome.[197] If this absence of PKSs and FASs from the Archaea holds up, perhaps they lost the potential for this aspect of metabolism after they became separate from the Bacteria and the Eucarya, or perhaps the last common ancestor of the three kingdoms had not yet evolved it, but did so after the Archaeal branch separated and before the Bacteria and Eucarya diverged.

On the hypothesis of a single common origin for typical FASs and PKSs, an early ancestor of present-day bacteria and eukaryotes might have evolved a primitive condensing enzyme that recruited other functions to become more efficient; addition of an acyl carrier protein and acyl transferases could have given rise to a rudimentary PKS, perhaps followed by recruitment of the reductive cycle to convert it to a FAS. The resulting primordial multifunctional synthase would have become further improved and diversified by subsequent mutation, recombination between diverged gene sequences, and gene duplication, opening the way for PKSs to evolve the ability to generate chemically distinct products, while the (by then) essential function of fatty acid biosynthesis could be retained by the organisms' FAS.

A type II structure is likely to have been primordial, with subsequent gene fusion generating the type I synthases. The finding that certain domains from the vertebrate type I FAS, separated from the multifunctional protein by proteolytic cleavage or genetic engineering, can function biochemically[169,198-200] is consistent with (although does not prove) an origin of the type I enzymes by domain fusion. Until the discovery of the type I bacterial synthases—the modular PKSs and the nonmodular FASs of *Brevibacterium*[13] and related bacteria[201]—the simplest hypothesis was that gene fusion accompanied the transition to eukaryotic cellular architecture. This perhaps reflected the need for coordinate, equimolar synthesis of catalytic sites in eukaryotes in the absence of the operon organization that, in bacteria, allows cotranscription and cotranslation of groups of adjacent genes. In this respect, the FASs and PKSs are not unusual: there are other examples of similar biochemistry being performed in bacteria by sets of discrete enzymes and in eukaryotes by single multifunctional enzymes.[202] Now it is a matter of more conjecture whether the proposed domain fusions actually occurred in a prokaryote and were inherited by a eukaryotic descendant; or in a eukaryote, after which they were transferred back to the ancestors of those present day bacteria that use type I synthases.

In any event, a minimum of two ancestral fusion cycles is needed to explain the fact that the type I synthases appear to form two families, based on domain order and nucleotide sequences: (1) the vertebrate FASs, the fungal PKSs, and each module of the bacterial modular PKSs; and (2) the fungal FASs and bacterial type I FASs (so far identified in *Brevibacterium* and *Mycobacterium*) in the other. The resemblance in FAS architecture between *Brevibacterium* and the fungi has recently been underlined by the unexpected finding of a second type I FAS in *Brevibacterium ammoniagenes* which not only resembles the fungal FASs in domain order, as does the first such enzyme to be discovered,[13] but even in carrying the site of a member of the newly discovered phosphopantetheinyltransferase superfamily to add the prosthetic group to the acyl carrier domain[203] in its C-terminus.[204] Resemblances between genes such as these, which appear to be at variance with taxonomic relationships, are highly suggestive of horizontal gene transfer. This term describes the hypothesis that genes may be acquired by one organism from another, not closely related to it, by a process other than their normal inheritance from an ancestor. Horizontal transfer is difficult to prove, but could be supported if differences emerged between phylogenetic trees based on FAS/PKS sequences and trees constructed from genes such as those for primary metabolic pathways.

We are on firmer ground in attributing the multimodular structure of the bacterial macrolide PKSs to repeated rounds of gene duplication, presumably within the bacteria themselves since there is (so far) no eukaryotic example. Support for this idea comes from the high degree of sequence resemblance at the DNA level between some of the modules. The current overall head-to-head/tail-to-tail models for the three-dimensional structure of the modular PKSs (section III.D) would accommodate very well a process of module addition, sometimes with gene fusion to generate covalent bonds between adjacent modules, and sometimes without, thus leaving a number of noncovalent "joints" in the assembly line. Synthases for increasingly complex polyketides could have evolved in this way. Subsequent exchange of modules could have contributed to the origin of a multitude of differently programmed PKSs from a limited number of components. As pointed out by Schwecke *et al.*,[147] the 14-module rapamycin PKS and the (presumed) 10-module FK506 PKS share an almost identically programmed four-module subunit (RapC and FkbA respectively), which would catalyze the last four condensations needed to build the rapamycin (42) and FK506 (43) polyketides (Figure 13); and Motamedi *et al.*[149] found that the sequences of some corresponding domains are more similar *between* RapC and FkbA than among modules in FkbA itself, suggesting that the two synthases may have acquired this subunit by a recent (horizontal) genetic exchange. Again, because the codon usage of the putative oleandomycin PKS gene of *S. antibioticus* is atypical for *Streptomyces*, it was suggested that this was a relatively recent acquisition A mosaic genetic architecture that evolved by the accretion, subtraction, and exchange of modules in various combinations could well underlie the striking observation that the large number of naturally occurring macrocyclic lactones and polyethers actually share

Table 5. Structures of Representative Polyketide and Fatty Acid Synthases

structural class of synthase	programming strategy	
	iterative	modular
type I	vertebrate FAS: one subunit fungal PKS: one subunit fungal FAS: two subunits *Brevibacterium* FAS: one subunit[a]	erythromycin PKS: six modules, three subunits rapamycin PKS: 14 modules, three subunits spiramycin PKS: seven modules, five subunits
type II	*E. coli* FAS: seven subunits actinorhodin PKS: six subunits[b]	none known

[a] But two kinds of FAS cooperate, at least for phosphopantetheinylation, and might perhaps associate noncovalently.[204] [b] Depends on definition: just 3 subunits (KS, CLF, ACP) for the minimal PKS.

a considerable degree of structural regularity, sometimes differing by a single chain-extender unit.[205]

B. Iterative vs Modular Programming

The most significant distinction between different types of organization among the FASs and PKSs is not whether the various catalytic sites that make up the toolkit of the synthase are covalently associated, as in the type I enzymes, or only noncovalently, as in the type II systems. In terms of programming mechanisms, the most relevant distinction is between synthases which carry a single set of sites that act iteratively in successive rounds of chain assembly and reduction, and those with a module of sites for each round, with each site acting just once in the building of an entire carbon chain. Table 5 exemplifies these distinctions. It reminds us that even in the type I systems, the complete synthase may consist of a series of different protein subunits (up to five among known examples) that presumably dock together to build the multifunctional enzyme. Presumably this docking must occur quite specifically in the case of the modular PKSs in order to establish the complete "assembly line", and the models of modular PKS structure allow for such head-to-tail docking. It is not obvious why a modular synthase should consist of any specific number of subunits (why for example the erythromycin and rapamycin PKSs are each built from three and the spiramycin PKS from five components); and after all seven monofunctional proteins can associate productively in the *E. coli* FAS. One *could* therefore imagine the existence of a type II modular PKS in which appropriate docking of domains, within as well as between modules, allowed it to function correctly, but no example has been discovered so far. A covalent structure for the modules is doubtless more efficient.

What was the selective pressure for the modular PKS to evolve? Two biochemical features currently correlate with the iterative vs modular construction of known FASs and PKSs. One is the use of exclusively malonyl CoA extender units, vs the use of at least a proportion of more complex chain extenders. The iterative synthases, including both the type I and the type II FASs, and the type I and type II synthases for the aromatic class of PKSs, use malonyl CoA for every round of chain extension (exceptions include FASs that introduce methyl branches,[1] but their pattern is quite repetitive and may not present a programming challenge), whereas the modular PKSs use either all methylmalonyl extenders (which yield two alternative stereochemical configurations for the incorporated propionate residues) as in the case of the erythromycin PKS (section IIID), or a precise sequence of malonyl, methylmalonyl, or more complex extenders (for example an ethylmalonyl unit to yield a butyrate residue in tylosin (**38**) and spiramycin (**39**), or glycerol-derived extenders in soraphen (**45**): Figure 13). Perhaps the choices necessitated by this variety of extender units were too hard to program into the iterative synthase, which is busy dealing with the issue of chain length that follows automatically from the number of modules in the modular PKSs. The second feature of the modular PKSs is the complexity of the programming of the reductive cycle that could generate five functionalities (keto, *R* hydroxy, *S* hydroxy, enoyl, or methylene) at each round of chain building, giving in principle 5^n possibilities, where n is the number of rounds (e.g., 15 625 for erythromycin, of which just one is chosen!). In contrast, aromatic PKSs usually make very simple reductive changes to the β-keto groups of the growing chain: e.g., none for resorcinol or tetracenomycin, or one reduction/dehydration for 6-methylsalicylic acid or actinorhodin. (And for examples such as actinorhodin there are even grounds to believe that the keto group to be reduced is actually recognized after chain assembly is complete, rather than reflecting a choice during chain building.[73]) Carrying out the complexity of reductive programming needed to build a macrolide aglycon may well be beyond the capacity of an iterative synthase.

Hopefully, a fuller understanding of the precise mechanisms of programming choices will eventually allow a distinction to be made between complexity of chain extender choices and of reductive cycle choices as a driving force in the evolution of the modular PKS. Meanwhile, what are needed are examples of PKSs that are required to make one or the other, but not both, kinds of choices. The polyketide chain of mevinolin (**59**: Figure 15), which is all acetate-derived and is made by an iterative synthase, contains double bonds and hydroxyl groups (discussed in ref 21), suggesting the need to program a significant number of reductive choices. Again, at least a part of the biosynthesis of coronafacic acid, the polyketide component of the *Pseudomonas* phytotoxin coronatine, is catalyzed by a type II (presumably iterative) PKS, but it is not at all certain that the product of this PKS includes the moiety of coronafacic acid that may need complex programming.[206] A further interesting case concerns pyoluteorin biosynthesis by *Pseudomonas fluorescens*.[206a] This is an aromatic polyketide assembled from all malonate CoA extender units, yet the PKS is modu-

lar. Perhaps this reflects the need for a ketoreduction associated with the third of these condensations.

C. Developmental Crosstalk

On the hypothesis of gene duplication and divergence in FAS and PKS evolution, many present day organisms will have come to contain two or more FASs and/or PKSs encoded by genes or gene clusters that have a common origin. These phylogenetically related synthases can be expected to retain varying degrees of functional and structural similarity. Do they also retain the potential for biochemical crosstalk, and if so are there situations in which this is exploited, perhaps as part of a developmental progression, or to maximize the metabolic diversity available from the appropriate expression of subsets of the structural genes? As the broader biology of fatty acid and polyketide synthesis is increasingly investigated, convincing evidence of productive crosstalk is indeed emerging. A few examples will serve as illustrations of what is becoming a fascinating field of study.

As described in section II, *Streptomyces coelicolor* contains two separate clusters of structurally similar genes that encode the type II PKSs for two different aromatic polyketides: the octaketide antibiotic actinorhodin and a (presumed) dodecaketide spore pigment. That at least some of the subunits of the two PKSs are (still) functionally compatible was shown by the artificial complementation of deletion or point mutations of the genes encoding any of the three minimal PKS subunits of either PKS by the corresponding subunits from the other.[78,109] Such crosstalk does not, however, occur naturally: mutations in the genes for one PKS are not compensated for by the unmutated genes for the other when the gene sets are expressed normally. Presumably this is because the normal expression pattern ensures that the actinorhodin PKS is made only in vegetative parts of the colony, while genes for the spore pigment PKS are expressed only in the developing spores. Probably there has been no selective pressure for the two PKSs to diverge far enough to avoid unwanted crosstalk because the opportunity for this does not arise. A different situation applies to the subunits of the type II FAS of *S. coelicolor*. Artificial coexpression of the FAS ACP (the other subunits have not been tested) with components of the actinorhodin PKS resulted in only a minute level of complementation: in this case, biochemical divergence has probably ensured the absence of unwanted crosstalk, which could have occurred because fatty acid biosynthesis may well still be needed in those parts of the colony that have started to make actinorhodin.[207] Interestingly, however, the gene for an essential component of any type II synthase, a malonyl transferase to transfer malonyl units from CoA to the prosthetic group of the ACP for chain extension, is found in the FAS gene cluster but in neither of the PKS clusters. In this case, conserved biochemical crosstalk may be essential to provide the two PKSs with a functional malonyl transferase,[208] and it is possible that fatty acid and polyketide synthases could even share the first condensation reaction.[207]

Another example concerns the fatty acid side chains of the lipooligosaccharide signaling molecules

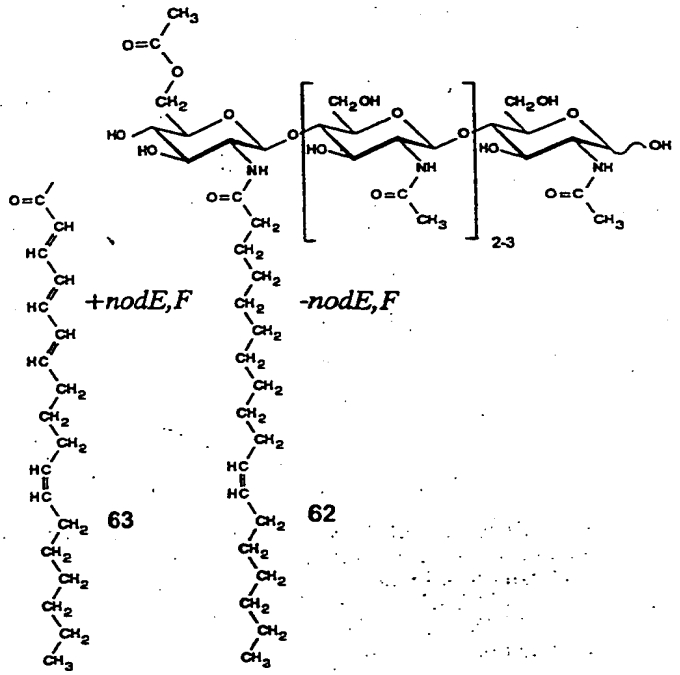


Figure 17. Structures of lipooligosaccharide nodulation factors produced by *Rhizobium leguminosarum* strains. In the absence of the *nodE*-encoded KS and *nodF*-encoded ACP, the acyl side chain is 18:1 characteristic of *cis*-vaccenic acid (**62**). When *nodE* and *nodF* are present, the 18:1 acyl side chain is replaced, in a proportion of the molecules, by the 18:4 unsaturated **63** to produce the highly bioactive Nod factor. (Modified from 209. Copyright 1991 Macmillan Magazines Ltd.)

produced by rhizobia. These molecules initiate the plant response that results in engulfment of the bacteria by the root hairs of leguminous plants and eventually leads to a nitrogen-fixing symbiosis within a specialized plant organ, the root nodule. The specificity of each signaling molecule for particular species of legume often depends on the length and pattern of unsaturation of the fatty acid side chains (e.g., **62** and **63**: Figure 17).[209] These appear to be made by a controlled kind of crosstalk between the type II FAS of the bacteria, which would synthesize a generic, largely saturated carbon chain up to at least C_{10}, and specific KS and ACP subunits encoded by nodulation genes, *nodE* and *nodF* (and in the case of *Rhizobium meliloti* also a KR encoded by *nodG*), which extend the chain in a specific manner, although presumably with the cooperation of other essential FAS subunits that are not provided *de novo* as *nod* gene products; and indeed there is an apparent competition between the primary FAS and the "nodulation FAS", so that a mixture of end products (**62** and **63**) is seen.[209]

There is a growing list of other situations in which fatty acid moieties form an integral part of specialized metabolites. Sometimes the evidence points to a gene duplication that has given rise to a complete special FAS for this role, distinct from the one involved in primary metabolism; a good example is the origin of the hexanoyl starter unit for norsolorinic acid biosynthesis in *Aspergillus* (see section IV).[185] Other situations may resemble the *Rhizobium* case, where the FAS of primary metabolism plays a second role by providing partially extended fatty acid chains that

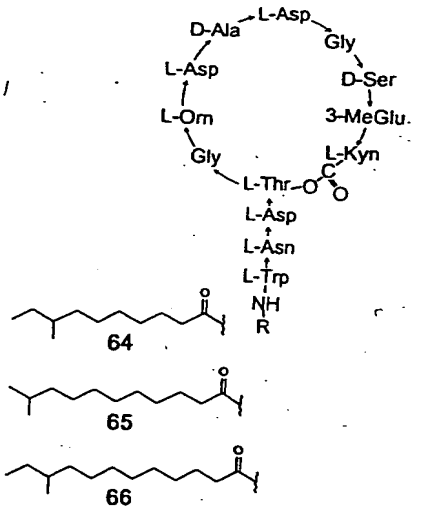


Figure 18. Structures of three members of the lipopeptide A21978C complex of *Streptomyces roseosporus*. The fatty acid side chains of **64** and **66** could start with 2-methylbutyryl CoA and **65** with isobutyl CoA or 3-methylbutyryl CoA, just as in the case of some of the major branched-chain fatty acids of the phospholipids of *Streptomyces*. (From ref 211.)

are "hijacked" for secondary metabolite biosynthesis, probably via the agency of the ACPs. An example is provided by the acyl side chains of the *N*-acyl homoserine lactones involved in quorum sensing in various Gram-negative bacteria, where an acyl ACP is believed to deliver the side chain to the enzyme that builds the lactone.[210] The fatty acid side chains of lipopeptide antibiotics such as the A21978C complex of *Streptomyces roseosporus*[111] may well originate in this way also, because of the very suggestive finding that three major components of the complex exist with C_{11}, C_{12} and C_{13} branched side chains (**63**–**65**: Figure 18); they correspond to truncated versions of some of the most abundant fatty acids of streptomycetes.

VII. Concluding Remarks

This article has been wide-ranging and has, I hope, given a flavor of the multidisciplinarity of current studies of polyketide synthesis. As a geneticist, I have emphasised (some might say overemphasized!) the contributions of genetics in its various forms—mutagenesis, sequence analysis, natural and engineered recombination, gene expression, and phylogenetic comparisons—to the development of knowledge in the field. Other contributors to this special issue of *Chemical Reviews* will, I am sure, focus much more on the chemistry and biochemistry of polyketide biosynthesis, and on the crucial part played by these disciplines in developing our understanding. It seems obvious that, in order to gain insight into the precise mechanisms of these astonishing multifunctional biosynthetic machines, *in vitro* studies will become preeminent over the next few years, although of course they will continue to be facilitated by protein engineering made possible by site-directed mutagenesis and the over-expression of genes. Genetic engineering will also be crucial in generating further recombinants for chemical anlysis, while of course *in vivo* studies will always be needed to help to discriminate between what is biochemically possible and what actually occurs in nature. And in the wider context of the biology of polyketide and fatty acid synthases genetics will surely continue to come up with novelties that will in turn stimulate further rounds of *in vitro* analysis.

VIII. Acknowledgments

I am very grateful to Mark Buttner, Tobias Kieser, Chaitan Khosla, Peter Revill, Joachim Schröder, Tom Simpson and Chandres Surti for providing figures, as detailed in the relevant captions, and to Mervyn Bibb, Keith Chater, Tobias Kieser, Peter Revill and Tom Simpson for helpful comments on the manuscript. Recent work in my laboratory has been supported by grants-in-aid from the Biotechnology and Biological Sciences Research Council (UK) and the John Innes Foundation, and project grants from the European Union (BIOT-CT91-0255, BI02-CT-94-2067 and CHRX-CT94-0570).

IX. References

(1) O'Hagan, D. *The Polyketide Metabolites*; Ellis Horwood: Chichester, 1991.
(2) Simpson, T. J. *Chem. Ind.* **1995**, 407.
(3) McCarthy, A. D.; Hardie, D. G. *Trends Biochem. Sci.* **1984**, *9*, 60.
(4) Magnuson, K.; Jackowski, S.; Rock, C. O.; Cronan, J. E. *Microbiol. Rev.* **1993**, *57*, 522.
(5) Amy, C. M.; Witkowski, A., Naggert, J.; Williams, B.; Rhandawa, Z.; Smith, S. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 3114.
(6) Holzer, K. P.; Liu, W.; Hammes, G. G. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4387.
(7) Schweizer, M.; Takabayashi, K.; Laux, T.; Beck, K. F. *Nucleic Acids Res.* **1989**, *17*, 567.
(8) Chirala, S. S.; Kuziora, M. A.; Spector, D. M.; Wakil, S. J. *J. Biol. Chem.* **1987**, *262*, 4231.
(9) Schweizer, M.; Roberts, L. M.; Höltke, H. J.; Takabayashi, K.; Höllerer, E.; Hoffmann, B.; Müller, G.; Köttig, H.; Schweizer, E. *Mol. Gen. Genet.* **1986** *202*, 479.
(10) Schweizer, E.; Müller, G.; Roberts, L. M.; Schweizer, M.; Rösch, J.; Wiesner, P.; Beck, J.; Stratmann, D.; Zauner, I. *Fat Sci. Technol.* **1987**, *89*, 570.
(11) Mohamed, A. H.; Chirala, S. S.; Mody, N. H.; Huang, W. Y.; Wakil, S. J. *J. Biol. Chem.* **1988**, *263*, 12315.
(12) Wiesner, P.; Beck, J.; Beck, K. F.; Ripka, S.; Müller, G.; Lücke, S.; Schweizer, E. *Eur. J. Biochem.* **1988**, *177*, 69.
(13) Meurer, G.; Biermann, G.; Schütz,; Harth, S.; Schweizer, E. *Mol. Gen. Genet.* **1992**, *232*, 106.
(14) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.
(15) Hopwood, D. A.; Sherman, D. H. *Annu. Rev. Genet.* **1990**, *24*, 37.
(16) Robinson, J. A. *Phil. Trans. R. Soc. Lond. B* **1991**, *332*, 107.
(17) Hopwood, D. A.; Khosla, C. In *Ciba Symposium 171: Secondary Metabolites: Their Function and Evolution*; Wiley: London, 1992; pp 88–112.
(18) O'Hagan, D. *Nat. Prod. Rep.* **1992**, *9*, 447.
(19) O'Hagan, D. *Nat. Prod. Rep.* **1993**, *10*, 593.
(20) Katz, L.; Donadio, S. *Annu. Rev. Microbiol.* **1993**, *47*, 875.
(21) Hutchinson, C. R.; Fujii, I. *Annu. Rev. Microbiol.* **1995**, *49*, 201.
(22) Richardson, M.; Khosla, C. In *Comprehensive Natural Products*; Elsevier: Oxford, in press.
(23) Bibb, M. J.; Schottel, J. L.; Cohen, S. N. *Nature* **1980**, *284*, 526.
(24) Thompson, C. J.; Ward, J. M.; Hopwood, D. A. *Nature* **1980**, *286*, 525.
(25) Suarez, J. E.; Chater, K. F. *Nature* **1980**, *286*, 527.
(26) Hopwood, D. A.; Bibb, M. J.; Bruton, C. J.; Chater, K. F.; Feitelson, J. S.; Gil, J. A. *Trends Biotechnol.* **1983**, *1*, 42.
(27) Hopwood, D. A. In *Biological, Biochemical and Biomedical Aspects of Actinomycetes*; Szabo, G., Biró, S., Goodfellow, M., Eds,; Akadémiai Kiadó: Budapest, 1986; pp 3–14.
(28) Chater, K. F.; Bruton, C. J. *EMBO J.* **1985**, *4*, 1893
(29) Malpartida, F.; Hopwood, D. A. *Nature* **1984**, *309*, 462.
(30) Motamedi, H.; Hutchinson, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4445.
(31) Butler, M. J.; Friend, E. J.; Hunter, I. S.; Kaczmarek, F. S.; Sugden, D. A.; Warren, M. *Mol. Gen. Genet.* **1989**, *215*, 231.
(32) Malpartida, F.; Hopwood, D. A. *Mol. Gen. Genet.* **1986**, *205*, 66.

(33) Malpartida, F.; Hallam, S. E.; Kieser, H. M.; Motamedi, H.; Hutchinson, C. R.; Butler, M. J.; Sugden, D. A.; Warren, M.; McKillop, C.; Bailey, C. R.; Humphreys, G. O.; Hopwood, D. A. *Nature* **1987**, *325*, 818.
(34) Arrowsmith, T. J.; Malpartida, F.; Sherman, D. H.; Birch, A.; Hopwood, D. A.; Robinson, J. A. *Mol. Gen. Genet.* **1992**, *234*, 254.
(35) Davis, N. K.; Chater, K. F. *Mol. Microbiol.* **1990**, *4*, 1679.
(36) Bergh, S.; Uhlén, M. *Gene* **1992**, *117*, 131.
(37) Hallam, S. E.; Malpartida, F.; Hopwood, D. A. *Gene* **1988**, *74*, 305.
(38) Fernández-Moreno, M. A.; Martínez, E.; Boto, L.; Hopwood, D. A.; Malpartida, F. *J. Biol. Chem.* **1992**, *267*, 19278.
(39) Hunter, I. S. Personal communication.
(40) Kim, E-S.; Hopwood, D. A.; Sherman, D. H. *Gene* **1994**, *141*, 141.
(41) Thamchaipenet, A. Ph.D. Thesis, University of Glasgow, 1994.
(42) Bibb, M. J.; Biró, S.; Motamedi, H.; Collins, J. F.; Hutchinson, C. R. *EMBO J.* **1989**, *8*, 2727.
(43) Summers, R. G.; Wendt-Pienkowski, E.; Motamedi, H.; Hutchinson, C. R. *J. Bacteriol.* **1992**, *174*, 1810.
(44) Summers, R. G.; Wendt-Pienkowski, E.; Motamedi, H.; Hutchinson, C. R. *J. Bacteriol.* **1993**, *175*, 7571.
(45) Sherman, D. H.; Malpartida, F.; Bibb, M. J.; Kieser, H. M.; Hopwood, D. A. *EMBO J.* **1989**, *8*, 2717.
(46) Stutzman-Engwall, K. J.; Hutchinson, C. R. *Proc. Natl. Acad. Aci. U.S.A.* **1989**, *86*, 3135.
(47) Grimm, A.; Madduri, K.; Ali, A.; Hutchinson, C. R. *Gene* **1994**, *151*, 1.
(48) Blanco, G.; Pereda, A.; Méndez, C.; Salas, J. A. *Gene* **1992**, *112*, 59.
(49) LeGouill, C.; Desmarais, D.; Déry, C. V. *Mol. Gen. Genet.* **1993**, *240*, 146.
(50) Bibb, M. J.; Sherman, D. H.; Ōmura, S.; Hopwood, D. A. *Gene* **1994**, *142*, 31.
(51) Yu, T-W.; Bibb, M. J.; Revill, W. P.; Hopwood, D. A. *J. Bacteriol.* **1994**, *176*, 2627.
(52) Han, L.; Yang, K.; Ramalingam, E.; Mosher, R. H.; Vining, L. C. *Microbiology* **1994**, *140*, 3379.
(53) Ye, J.; Dickens, M. L.; Plater, R.; Li, Y.; Lawrence, J.; Strohl, W. R. *J. Bacteriol.* **1994**, *176*, 6270.
(54) Piecq, M.; Dehottay, P.; Biot, A.; Dusart, J. *J. DNA Sequencing Mapping* **1994**, *4*, 219.
(55) Decker, H.; Haag, S. *J. Bacteriol.* **1995**, *177*, 21.
(56) Ylihonko, K.; Tuikkanen, J.; Jussila, S.; Cong, L.; Mäntsälä, P. *Mol. Gen. Genet.* **1996**, *251*, 113.
(57) Ylihonko, K.; Hakala, J.; Kunnari, T.; Mäntsälä, P. *Microbiology* **1996**, *142*, 1965.
(58) Lombó, F.; Blanco, G.; Fernández, E.; Méndez, C.; Salas, J. A. *Gene* **1996**, *172*, 87.
(59) Niemi, J.; Ylihonko, K.; Hakala, J.; Pärssinen, R.; Kopio, A.; Mäntsälä, P. *Microbiology* **1994**, *140*, 1351.
(60) Tsukamoto, N.; Fujii, I.; Ebizuka, Y.; Sankawa, U. *J. Bacteriol.* **1994**, *176*, 2473.
(61) Kakinuma, S.; Ikeda, H.; Takeda, Y.; Tanaka, H.; Hopwood, D. A.; Ōmura, S. *J. Antibiot.* **1995**, *48*, 484.
(62) Decker, H.; Rohr, J.; Motamedi, H.; Zähner, H.; Hutchinson, C. R. *Gene* **1995**, *166*, 121.
(63) Hong, S-T.; Carney, J. R.; Gould, S. J. *J. Bacteriol.* **1997**, *179*, 470.
(64) Rudd, B. A. M.; Hopwood, D. A. *J. Gen. Microbiol.* **1979**, *114*, 35.
(65) Kauppinen, S.; Siggaard-Andersen, M.; von Wettstein-Knowles, P. *Carlsberg Res. Commun.* **1988**, *53*, 357.
(66) Zalkin, H.; Ebbole, D. J. *J. Biol. Chem.* **1988**, *263*, 1595.
(67) Sherman, D. H.; Kim, E.-S.; Bibb, M. J.; Hopwood, D. A. *J. Bacteriol.* **1992**, *174*, 6184.
(68) Zhang, H. L.; He, X. G.; Adefarati, A.; Gallucci, J.; Cole, S. P.; Beale, J. M.; Keller, P. J.; Chang, C. J.; Floss, H. G. *J. Org. Chem.* **1993**, *55*, 1682.
(69) Sherman, D. H.; Bibb, M. J.; Simpson, T. J.; Johnson, D.; Malpartida, F.; Fernández-Moreno, M.; Martínez, E.; Hutchinson, C. R.; Hopwood, D. A. *Tetrahedron* **1991**, *47*, 6029.
(70) Hopwood, D. A.; Malpartida, F.; Kieser, H. M.; Ikeda, H.; Duncan, J.; Fujii, I.; Rudd, B. A. M.; Floss, H. G.; Ōmura, S. *Nature* **1985**, *314*, 642.
(71) Ōmura, S.; Ikeda, H.; Malpartida, F.; Kieser, H. M.; Hopwood, D. A. *Antimicrob. Agents Chemother.* **1986**, *29*, 13.
(72) Epp, J. K.; Huber, M. L. B.; Turner, J. R.; Goodson, T.; Schoner, B. E. *Gene* **1989**, *85*, 293.
(73) Bartel, P. L.; Zhu, C. B.; Lampel, J. S.; Dosch, D. C.; Connors, N. C.; Strohl, W. R.; Beale, J. M.; Floss, H. G. *J. Bacteriol.* **1990**, *172*, 4816.
(74) Gorst-Allman, C. P.; Rudd, B. A. M.; Chang, C.-J.; Floss, H. G. *J. Org. Chem.* **1981**, *46*, 455.
(75) Floss, H. G.; Cole, S. P.; He, X.-G.; Rudd, B. A. M.; Duncan, J.; Fujii, I.; Chang, C.-J.; Keller, P. J. In *Regulation of Secondary Metabolite Formation;* Kleinkauf, H., von Döhren, H., Dornauer, H., Nesemann, G., Eds.; VCH: Weinheim, 1986; pp 283–304.
(76) Khosla, C.; Ebert-Khosla, S.; Hopwood, D. A. *Mol. Microbiol.* **1992**, *6*, 3237.
(77) Khosla, C.; McDaniel, R.; Ebert-Khosla, S.; Torres, R.; Sherman, D. H.; Bibb, M. J.; Hopwood, D. A. *J. Bacteriol.* **1993**, *175*, 2197.
(78) Kim, E.-S.; Hopwood, D. A.; Sherman, D. H. *J. Bacteriol.* **1994**, *176*, 1801.
(79) Hopwood, D. A.; Khosla, C.; Sherman, D. H.; Bibb, M. J.; Ebert-Khosla, S.; Kim, E.-S.; McDaniel, R.; Revill, W. P.; Torres, R.; Yu, T.-W. In *Industrial Microorganisms: Basic and Applied Molecular Genetics*; Baltz, R. H., Hegeman, G. D., Skatrud, P. L., Eds.; American Society for Microbiology, Washington, 1993; pp 267–275.
(80) Hopwood, D. A. *Curr. Opin. Biotechnol.* **1993**, *4*, 531.
(81) Fernández-Moreno, M. A.; Caballero, J. L.; Hopwood, D. A.; Malpartida, F. *Cell* **1991**, *66*, 769.
(82) Fernández-Moreno, M. A.; Martínez, E.; Caballero, J. L.; Ichinose, K.; Hopwood, D. A.; Malpartida, F. *J. Biol. Chem.* **1994**, *269*, 24854.
(83) Caballero, J. L.; Martínez, E.; Malpartida, F.; Hopwood, D. A. *Mol. Gen. Genet.* **1991**, *230*, 401.
(84) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *Science* **1993**, *262*, 1546.
(85) Hopwood, D. A.; Bibb, M. J.; Chater, K. F.; Kieser, T.; Bruton, C. J.; Kieser, H. M.; Lydiate, D. J.; Smith, C. P.; Ward, J. M.; Schrempf, H. In *Genetic Manipulation of Streptomyces. A laboratory manual*: John Innes Foundation: Norwich, 1985.
(86) Lydiate, D. J.; Malpartida, F.; Hopwood, D. A. *Gene* **1985**, *35*, 223.
(87) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1994**, *116*, 10855.
(88) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11542.
(89) Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1994**, *116*, 4166.
(90) Meurer, G.; Hutchinson, C. R. *J. Bacteriol.* **1995**, *177*, 477.
(91) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1993**, *115*, 11671.
(92) Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *Biochemistry* **1994**, *33*, 9321.
(93) Fu, H.; Hopwood, D. A.; Khosla, C. *Chem. Biol.* **1994**, *1*, 205.
(94) McDaniel, R.; Hutchinson, C. R.; Khosla, C. *J. Am. Chem. Soc.* **1995**, *117*,
(95) Alvarez, M. A.; Fu, H.; Khosla, C.; Hopwood, D. A.; Bailey, J. E. *Nature Biotechnol.* **1996**, *14*, 335.
(96) Kramer, P. J.; Zawada, R. J. X.; McDaniel, R.; Hutchinson, C. R.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1997**, *119*, 635.
(97) Yu, T.-W. Ph.D. Thesis, University of East Anglia, Norwich, 1995.
(98) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *Nature* **1995**, *375*, 549.
(99) Tsoi, C. J.; Khosla, C. *Chem. Biol.* **1995**, *2*, 355.
(100) Khosla, C.; Zawada, R. J. X. *Trends Biotechnol.* **1996**, *14*, 335.
(101) Garwin, J. L.; Klages, A. L.; Cronan, J. E., Jr. *J. Biol. Chem.* **1980**, *255*, 11949. (a) Summers, R. G.; Ali, A.; Shen, B.; Hutchinson, C. R. *J. Bacteriol.* **1995**, *34*, 9389.
(102) Rohr, J. *J. Org. Chem.* **1992**, *57*, 5217.
(103) Thomas, R. *Folia Microbiol.* **1995**, *40*, 4.
(104) Zawada, R. J. X.; Khosla, C. *J. Biol. Chem.* **1997**, *272*, 16184.
(105) Yu, T.-W.; McDaniel, R.; Moore, B.; Shen, S.; Khosla, C.; Hopwood, D. A. Manuscript in preparation, 1997.
(106) Thomas, R.; Williams, D. J. *J. Chem. Soc., Chem. Commun.* **1983**, *667*.
(107) Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1994**, *116*, 6443.
(108) Rajgarhia, V. B.; Strohl, W. R. *J. Bacteriol.* **1997**, *179*, 2690.
(109) Yu, T.-W.; Hopwood, D. A. *Microbiology* **1995**, *141*, 2779.
(110) Blanco, G.; Fu, H.; Mendez, C.; Khosla, C.; Salas, J. A. *Chem. Biol.* **1996**, *3*, 193.
(111) Griffin, D. A.; Leeper, F. J.; Staunton, J. *J. Chem. Soc., Perkin Trans.* **1984**, *1*, 1035.
(112) Gramajo, H. C.; White, J.; Hutchinson, C. R.; Bibb, M. J. *J. Bacteriol.* **1991**, *173*, 6475.
(113) Shen, B.; Summers, R. G.; Gramajo, H.; Bibb, M. J.; Hutchinson, C. R. *J. Bacteriol.* **1992**, *174*, 3818.
(114) Crosby, J.; Sherman, D. H.; Bibb, M. J.; Revill, W. P.; Hopwood, D. A.; Simpson, T. J. *Biochim. Biophys. Acta* **1995**, *1251*, 32.
(115) Crump, M. P.; Crosby, J.; Dempsey, C. E.; Murray, M.; Hopwood, D. A.; Simpson, T. J. *FEBS Lett.* **1996**, *391*, 302.
(116) Crump, M. P.; Crosby, J.; Dempsey, C. E.; Parkinson, J. A.; Murray, M.; Hopwood, D. A.; Simpson, T. J. *Biochemistry* **1997**, *36*, 6000.
(117) Shen, B.; Hutchinson, C. R. *Science* **1993**, *262*, 1535.
(118) Shen, B.; Hutchinson, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6600.
(119) Carreras, C. W.; Pieper, R.; Khosla, C. *J. Am. Chem. Soc.* **1996**, *118*, 5158.
(120) Thompson, C. J.; Ward, J. M.; Hopwood, D. A. *J. Bacteriol.* **1982**, *151*, 668.
(121) Bibb, M. J.; Janssen, G. R.; Ward, J. M. *Gene* **1985**, *1405*, E357.
(122) Uchiyama, H.; Weisblum, B. *Gene* **1985**, *38*, 103.

(123) Stanzak, R.; Matsushima, P.; Baltz, R. H.; Rao, R. N. *Bio/Technology* **1986**, *4*, 229.
(124) Dhillon, N.; Hale, R. S.; Cortes, J.; Leadlay, P. F. *Mol. Microbiol.* **1989**, *3*, 1405.
(125) Vara, J.; Lewandowska-Skarbek, M.; Wang, Y.-G.; Donadio, S.; Hutchinson, C. R. *J. Bacteriol.* **1989**, *171* 5872.
(126) Paulus, T. J.; Tuan, J. S.; Luebke, V. E.; Main, G. T.; DeWitt, J. P.; Katz, L. *J. Bacteriol.* **1990**, *172*, 2541.
(127) Weber, M. J.; Leung, J. O.; Maine, G. T.; Potenz, R. H. B.; Paulus, T. J.; DeWitt, J. P. *J. Bacteriol.* **1990**, *172*, 2372.
(128) Haydock, S. F.; Dowson, J. A.; Dhillon, N.; Roberts, G. A.; Cortes, J.; Leadlay, P. F. *Mol. Gen. Genet.* **1991**, *230*, 120.
(129) Tuan, J. S.; Weber, J. M.; Staver, M. J.; Leung, J. O.; Donadio, S.; Katz, L. *Gene* **1990**, *90*, 21.
(130) Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F. *Nature* **1990**, *348*, 176.
(131) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. *Science* **1991**, *252*, 675.
(132) Bevitt, D. J.; Cortes, J.; Haydock, S. F.; Leadlay, P. F. *Eur. J. Biochem.* **1992**, *204*, 39.
(133) Caffrey, P.; Bevitt, D. J.; Staunton, J.; Leadlay, P. F. *FEBS Lett.* **1992**, *304*, 225.
(134) Baltz, R. H.; Seno, E. T. *Antimicrob. Agents Chemother.* **1981**, *20*, 214.
(135) Baltz, R. H. In *Genetic Engineering of Microorganisms for Chemicals;* Hollaender, A., DeMoss, R. D., Kaplan, S., Konisky, J., Savage, D., Wolfe, R. S., Eds.; Plenum Publishing Corporation: New York, 1982; pp 431–444.
(136) Fishman, S. E.; Cox, K.; Larson, J. L.; Reynolds, P. A.; Seno, E. T.; Yeh, W.-K.; van Frank, R.; Hershberger, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8248.
(137) Beckmann, R. J.; Cox, K.; Seno, E. T. In *Genetics and Molecular Biology of Industrial Microorganisms;* Hershberger, C. L., Queener, S. W., Hegeman, G., Eds.; American Society for Microbiology: Washington, DC, 1989; pp 176–186.
(138) Fishman, S. E.; Rosteck, P. R., Jr.; Hershberger, C. L. *J. Bacteriol.* **1985**, *161*, 199.
(139) Young, M.; Cullum, J. *FEMS Microbiol. Lett.* **1987**, *212*, 10.
(140) Kuhstoss, S.; Huber, M.; Turner, J. R.; Paschal, J. W.; Rao, R. N. *Gene* **1996**, *183*, 231.
(141) Richardson, M. A.; Kuhstoss, S.; Huber, M.; Ford, L.; Godfrey, O.; Turner, J. R.; Rao, R. N. In *Genetics and Molecular Biology of Industrial Microorganisms*; Hershberger, C. L., Queener, S. W., Hegeman, G., Eds.; American Society for Microbiology: Washington, DC, 1989; pp 40–43.
(142) Epp, J. K.; Huber, M. L.; Turner, J. R.; Schoner, B. E. In *Genetics and Molecular Biology of Industrial Microorganisms*; Hershberger, C. L., Queener, S. W., Hegeman, G., Eds.; American Society for Microbiology: Washington, DC, 1989; pp 35–39.
(143) Swan, D. G.; Rodríguez, A. M.; Vilches, C.; Méndez, C.; Salas, J. A. *Mol. Gen. Genet.* **1994**, *242*, 358.
(144) Streicher, S. L.; Ruby, C. L.; Paress, P. S.; Sweasy, J. B.; Danis, S. J.; MacNeil, D.; Gewain, K.; MacNeil, T.; Foor, F.; Morin, N.; Cimis, G.; Rubin, R.; Goldberg, R. B.; Nallin, M.; Schulman, M. D.; Gibbons, P. In *Genetics and Molecular Biology of Industrial Microorganisms*; Hershberger, C. L., Queener, S. W., Hegeman, G., Eds.; American Society for Microbiology: Washington, DC, 1989; pp 44–52.
(145) MacNeil, D. J.; Occi, J. L.; Gewain, K. M.; MacNeil, T.; Gibbons, P. H.; Ruby, C. L.; Danis, S. J. *Gene* **1992**, *115*, 119.
(146) MacNeil, D. J.; Occi, J. L.; Gewain, K. M.; MacNeil, T.; Gibbons, P. H.; Foor, F.; Morin, N. In *Industrial Microorganisms: Basic and Applied Molecular Genetics*; Baltz, R. H., Hegeman, G. D., Skatrud, P. L., Eds.; American Society for Microbiology: Washington, DC, 1993; pp 245–256.
(147) Schwecke, T.; Aparicio, J. F.; Molnar, I.; Konig, A.; Khaw, L. E.; Haydock, S. F. Oliynyk, M.; Caffrey, P.; Cortes, J.; Lester, J. B.; Bohm, G. A.; Staunton, J.; Leadlay, P. F. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7839. (a) Lomovskaya, N.; Fonstein, L.; Ruan, X.; Stassi, D.; Katz, L.; Hutchinson, C. R. *Microbiology* **1987**, *143*, 875–883.
(148) Motamedi, H.; Shafiee, A.; Cai, S.-J.; Streicher, S. L.; Arison, B. H.; Miller, R. R. *J. Bacteriol.* **1996**, *178*, 5243.
(149) Motamedi, H.; Cai, S.-J.; Shafiee, A.; Elliston, K. O. *Eur. J. Biochem.* **1997**, in press.
(150) Hu, Z.; Bao, K.; Zhou, X.; Zhou, Q.; Hopwood, D. A.; Kieser, T.; Deng, Z. *Mol. Microbiol.* **1994**, *14*, 163.
(151) Schupp, T.; Toupet, C.; Cluzel, B.; Neff, S.; Hill, S.; Beck, J. J.; Ligon, J. M. *J. Bacteriol.* **1995**, *177*, 3673.
(152) Donadio, S.; McAlpine, J. B.; Sheldon, P. J.; Jackson, M.; Katz, L. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7119.
(153) Kao, C. M.; Katz, L.; Khosla, C. *Science* **1994**, *265*, 509.
(154) Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1994**, *116*, 11612.
(155) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. *Gene* **1992**, *115*, 97.
(156) Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla. C. *J. Am. Chem. Soc.* **1995**, *117*, 9105.
(157) Cortes, J.; Wiesmann, K. E. H.; Roberts, G. A.; Brown, M. J. B.; Staunton, J.; Leadlay, P. F. *Science* **1995**, *268*, 1487.
(158) Brown, M. J. B.; Cortes, J.; Cutter, A. L.; Leadlay, P. F.; Staunton, J. *J. Chem. Soc., Chem. Commun.* **1995**, 1517.
(159) Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1996**, *118*, 9184.
(160) Oliynyk, M.; Brown, M. J. B.; Cortés, J.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **1996**, *3*, 833.
(161) Leadlay, P. F.; Staunton, J. Personal communication. (a) McDaniel, R.; Kao, M. C.; Fu, H.; Hevezi, P.; Gustafsson, K.; Betlach, M.; Ashley, G.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1997**, *119*, 4309.
(162) Marsden, A. F. A.; Caffrey, P.; Aparicio, J. F.; Loughran, M. S.; Staunton, J.; Leadlay, P. F. *Science* **1994**, *263*, 378.
(163) Pieper, R.; Luo, G.; Cane, D. E.; Khosla, C. *Nature* **1995**, *378*, 263.
(164) Wiesmann, K. E. H.; Cortés, J.; Brown, M. J. B.; Cutter, A. L.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **1995**, *2*, 583.
(165) Pieper, R.; Luo, G.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1995**, *117*, 11373.
(166) Pieper, R.; Ebert-Khosla, S.; Cane, D.; Khosla, C. *Biochemistry* **1996**, *35*, 2054.
(167) Staunton, J.; Caffrey, P.; Aparicio, J. F.; Roberts, G. A.; Bethell, S. S.; Leadlay, P. F. *Nature Struct. Biol.* **1996**, *3*, 188.
(168) Kao, C. M.; Pieper, R.; Cane, D. E.; Khosla, C. *Biochemistry* **1996**, *35*, 12363.
(169) Wakil, S. J. *Biochemistry* **1989**, *28*, 4523.
(170) Smith, S. *FASEB J.* **1997**, *8*, 1248.
(171) Dimroth, P.; Walter, H.; Lynen, F. *Eur. J. Biochem.* **1970**, *13*, 98.
(172) Beck, J.; Ripka, S.; Siegner, A.; Schiltz, E.; Schweizer, E. *Eur. J. Biochem.* **1990**, *192*, 487.
(173) Schorr, R.; Mittag, M.; Müller, G.; Schweizer, E. *J. Plant Physiol.* **1994**, *143*, 407.
(174) Bedford, D. J.; Schweizer, E.; Hopwood, D. A.; Khosla, C. *J. Bacteriol.* **1995**, *177*, 4544.
(175) Mayorga, M. E.; Timberlake, W. E. *Mol. Gen. Genet.* **1992**, *235*, 205.
(176) Chang, P.-K.; Cary, J. W.; Yu, J.; Bhatnagar, D.; Cleveland, T. E. *Mol. Gen. Genet.* **1995**, *248*, 270.
(177) Feng, G. H.; Leonard, T. J. *J. Bacteriol.* **1995**, *177*, 6246.
(178) Trail, F.; Mahanti, N.; Rarick, M.; Mehigh, R.; Liang, S.-H.; Zhou, R.; Linz, J. E. *Appl. Environ. Microbiol.* **1995**, *61*, 2665.
(179) Kimura, N.; Tsuge, T. *J. Bacteriol.* **1995**, *175*, 4427.
(180) Takano, Y.; Kubo, Y.; Shimizu, K.; Mise, K.; Okuno, T.; Furusawa, I. *Mol. Gen. Genet.* **1995**, *249*, 162.
(181) International Patent no. WO 95/12661.
(182) Brown, D. W.; Yu, J.-H.; Kelkar, H. S.; Fernandes, M.; Nesbitt, T. C.; Keller, N. P.; Adams, T. H.; Leonards, T. J. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1418.
(183) Mahanti, N.; Bhatnagar, D.; Cary, J. W.; Joubran, J.; Linz, J. E. *Appl. Environ. Microbiol.* **1996**, *62*, 191.
(184) Watanabe, C. M. H.; Wilson, D.; Linz, J. E.; Townsend, C. A. *Chem. Biol.* **1996**, *3*, 463.
(185) Brown, D. W.; Adams, T. H.; Keller, N. P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14873.
(186) Townsend, C. A.; McGuire, S. M.; Brobst, S. W.; Graybill, T. L.; Pal, K.; Barry, C. E. In *Secondary-metabolite biosynthesis and metabolism*; Petroski, R. J., McCormick, S. P., Eds.; Plenum Press: New York, 1991; pp 141–154.
(187) Lanz, T.; Tropf, S.; Marner, F.-J.; Schröder, J.; Schröder, G. *J. Biol. Chem.* **1991**, *266*, 9971. (a) Schröder, J. In *Comprehensive Natural Product Chemistry*; Elsevier: Oxford, Vol. 1, in press.
(188) Reimold, U.; Kröger, M.; Kreuzaler, F.; Hahlbrock, K. *EMBO J.* **1983**, *2*, 1801.
(189) Schröder, J.; Schröder, G. *Z. Naturforsch.* **1989**, *45c*, 1.
(190) Tropf, S.; Lanz, T.; Rensing, S. A.; Schröder, J.; Schröder, G. *J. Mol. Evol.* **1994**, *38*, 610.
(191) Zuurbier, K. W. M.; Fung, S.-Y.; Scheffer, J. J. C.; Verpoorte, R. *Phytochemistry* **1995**, *38*, 77.
(192) Tropf, S.; Kärcher, B.; Schröder, G.; Schröder, J. *J. Biol. Chem.* **1995**, *270*, 7922.
(193) Martin, C. R. *Int. Rev. Cytol.* **1993**, *147*, 233.
(194) Spaink, H. P. *Overdruk van de Bekroonde Inzending voor de Kluyver Prijs* **1993**.
(195) Siggaard-Andersen, M. *Protein Sequences Data Anal.* **1993**, *5*, 325.
(196) Revill, W. P. Personal communication.
(197) Bult, C. J., *et al. Science* **1996**, *273*, 1058.
(198) Naggert, J.; Witkowski, A.; Wessa, B.; Smith, S. *Biochem. J.* **1991**, *273*, 787.
(199) Rangan, V. S.; Smith, S. *J. Biol. Chem.* **1996**, *271*, 31749.
(200) Tropf, S.; Revill, W. P.; Bibb, M. J.; Hopwood, D. A.; Schweizer, M. Manuscript in preparation, 1997.
(201) Fernandes, N. D.; Kolattakudy, P. E. *Gene* **1996**, *170*, 95.
(202) Lumsden, J.; Coggins, J. *Biochem. J.* **1978**, *169*, 441.
(203) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. *Chem. Biol.* **1996**, *3*, 923.
(204) Stuible, H.-P.; Meurer, G.; Wagner, C.; Huter, G.; Schweizer, E. Abstracts, Symposium on the Enzymology of Biosynthesis of Natural Products, T. U. Berlin, 22–25 September 1996, p 25.

(205) Celmer, W. D. *Ann. N.Y. Acad. Sci.* **1986**, *471*, 299 and references therein.
(206) Penfold, C. N.; Bender, C. L.; Turner, J. G. *Gene* **1996**, *183*, 167. (a) Novak-Thompson, B.; Gould, S. J.; Loper, J. E. *Gene* **1997**, in press.
(207) Revill, W. P.; Bibb, M. J.; Hopwood, D. A. *J. Bacteriol.* **1996**, *178*, 5660.
(208) (a) Revill, W. P.; Bibb, M. J.; Hopwood, D. A. *J. Bacteriol.* **1995**, *177*, 3946. (b) Summers, R. G.; Ali, A.; Shen, B.; Wessel, A.; Hutchinson, C. R. *Biochemistry* **1995**, *34*, 9389.
(209) Spaink, H. P.; Sheeley, D. M.; van Brussel, A. A. N.; Glushka, J.; York, W. S.; Tak, T.; Geiger, O.; Kennedy, E. P.; Reinhold, V. N.; Lugtenberg, B. J. J. *Nature* **1991**, *354*, 125.
(210) Swift, S.; Stewart, G. S. A. B.; Williams, P. *Trends Microbiol.* **1996**, *4*, 463.
(211) Debono, M.; Barnhart, M.; Carrell, C. B.; Hoffmann, J. A.; Occolowitz, J. L.; Abbott, B. J.; Fukuda, D. S.; Hamill, R. L.; Biemann, K.; Herlihy, W. C. *J. Antibiot.* **1986**, *40*, 761.

CR960034I